Topics in Classical and Quantum Linear Stochastic Systems

Hendra Ishwara Nurdin

April 2007

A thesis submitted for the degree of Doctor of Philosophy
of the Australian National University

Dedicated to open minded and peace loving people everywhere

Declaration

I hereby declare that the work in this thesis are my own, or partially in collaboration with others, except where otherwise stated and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which to a substantial extent has been accepted for the award of any other degree or diploma of the university or other institute of higher learning. My contributions to results of collaborative work presented in this thesis are at least 85%.

Hendra I. Nurdin

List of Publications

This thesis is based on the following collection of papers which have either been published in or submitted to peer-reviewed journals or conferences:

I. Journal publications

1. M. R. James, H. I. Nurdin and I. R. Petersen, "H^∞ control of linear quantum systems," submitted to *IEEE Transactions on Automatic Control*, 2006, preprint: arxiv.org/quant-ph/0703150 (Chapters 7 and 8).

2. H. I. Nurdin, "Spectral factorization of a class of matrix-valued spectral densities", *SIAM Journal on Control and Optimization*, 45(5), pp. 1801-1821, 2006 (Chapter 6).

3. H. I. Nurdin, "New results on the rational covariance extension problem with degree constraint", *Systems and Control Letters*, 55(7), pp. 530-537, July 2006 (Chapter 4).

4. H. I. Nurdin and A. Bagchi, "On the solutions of the rational covariance extension problem corresponding to psedopolynomials having boundary zeros", *IEEE Transactions on Automatic Control*, 51(2), pp. 350-355, February 2006 (Chapter 3).

II. Conference publications

1. H. I. Nurdin and J. B. Moore, "Computation of degree constrained rational interpolants with non-strictly positive parametrizing functions via homotopy continuation," in Proceedings of the 45th IEEE Conference on Decision and Control (San Diego, CA, USA, December 13-15, 2006), pp. 565-570, 2006 (Chapter 5).

2. H. I. Nurdin, "A new approach to spectral factorization of a class of matrix-valued spectral densities," in Proceedings of the Joint 44th IEEE conference

on Decision and Control (CDC) and European Control Conference (ECC) (Seville, Spain, December 12-15, 2005), pp. 5929-5934, 2005 (Chapter 6).

3. H. I. Nurdin and A. Bagchi, "On the solutions of the rational covariance extension problem corresponding to psedopolynomials having boundary zeros," in Proceedings of the 43rd IEEE Conference on Decision and Control (Atlantis, Paradise Island, The Bahamas, December 14-17, 2004), pp. 5386-5391, December 2004 (Chapter 3).

The journal paper (4) and conference paper (3), and the journal paper (2) and conference paper (2), contain some overlapping materials.

Acknowledgements

This thesis would not have been possible without the many people who guided and supported me along the way. First and foremost, I am greatly indebted to my supervisors Prof. John B. Moore and Prof. Matthew R. James. John has been incredibly supportive from even before my arrival at ANU and remained so throughout my study. He gave me a lot of freedom to pursue my interests and to try to connect with different people, but at the same time was constructively critical of various aspects of my study and professional development. I have learned a lot from John and it has been my privilege to be his last PhD student. I am grateful to Matt for taking an interest in me and giving me an opportunity to work with him in the emerging and exciting world of quantum control. Matt has been very patient in guiding me as I started out with zero knowledge of quantum mechanics, he suggested the right materials to read and allowed me to pick them up at my own pace. I have found quantum control to be a fascinating subject and have immensely enjoyed the learning experience and collaboration with Matt.

My thanks also goes to Prof. Arunabha Bagchi who supported me to do research with him at the University of Twente, and to Prof. Tryphon Georgiou for helpful discussions on some topics in rational interpolation and spectral factorization. During my study I have also benefited from interactions and discussions with other members of the control and physics community and in this respect I wish to thank Prof. Ian Petersen, Dr. John Gough, Prof. Ravi Mazumdar, Dr. Valery Ougrinovskii and Dr. Subhrakanti Dey.

From December 1-8, 2006 I was fortunate to have had an opportunity to visit the Physical Measurement and Control group (read: MabuchiLab) at the California Institute of Technology and would like to thank them for their hospitality, with special thanks to Dr. Luc Bouten and Dr. Ramon van Handel for enlightening discussions and their constructive feedback on the work which appears in Part II of this thesis, and to Dr. Naoki Yamamoto for an interesting discussion on quantum relative entropy. I am also grateful to Luc for tirelessly giving me

an invaluable intensive tutorial on quantum filtering over a number of long days at CalTech.

On the social side, I thank various colleagues at the RSISE and FEIT, especially my officemates Desmond, Arvin, Andrew and Huibo in RSISE B-345, for their friendship and for ... the soccer games (at least up until November 2006 when I called it quits) which, once upon a time, was the social highlight of my week. Special mention to Desmond Chik and Daniel Heckenberg, a colleague from UNSW and frequent visitor to ANU, for volunteering themselves to be my "guinea pigs" for a private, informal experimental crash course I ran on selected topics in analysis and differential geometry which are of interest in engineering. Thanks also to my former colleagues in the Systems, Signals and Control group at the applied mathematics department of the University of Twente, whom have now gone their separate ways, for the close friendship we shared and the many fun things we use to do together. Special thanks to Kam for being a good friend and for her willingness to always lend a hand when I need it.

I also gratefully acknowledge the generous financial support I have received from ANU and NICTA for my PhD study and research.

Finally, unbounded thanks to Mom Linda, Dad Hazli and my siblings Rahmi, Yudi, Ardi, Nuzul and Ika for their enduring support and love without which I would not have made it this far.

Abstract

This thesis considers two topics in the area of linear stochastic systems.

The first topic is the construction of approximate finite dimensional linear time invariant (LTI) models for classical wide sense stationary stochastic signals with a non-coercive and non-rational spectral density, utilizing the recently developed theory of degree constrained rational interpolation. Non-coercive means that the spectral density has zeros on the unit circle or the imaginary axis (depending on whether the stochastic process is in discrete or continuous time, respectively), while non-rationality implies that the underlying system generating such a signal is infinite dimensional. As one example, spectral densities of this type appear when measurements are taken of signals that have traversed through the earth's turbulent atmosphere, such as light from a distant star captured by astronomical telescopes on the ground. The operation of obtaining an LTI model from a spectral density is known in the literature as *spectral factorization* and has played an important role in both deterministic and stochastic linear systems theory. The non-rational and non-coercive spectral densities which are considered herein are known to be difficult to factorize numerically. The most general algorithms for spectral factorization, such as the maximum entropy method, converge slowly for these spectral densities, and can lead to approximate models of degree higher than is necessary. The first part of this thesis establishes some new results in the theory of degree constrained rational interpolation and then proposes and analyzes a new approach to spectral factorization, based on so-called rational covariance extensions. A new spectral factorization algorithm is then introduced. In a number of simulations, which include some physically motivated spectral densities, it is demonstrated that the new algorithm gives lower degree approximations than the well-known maximum entropy method.

The second topic is the issue of physical realizability of a given system represented by linear quantum stochastic differential equations (QSDEs) on a quantum probability space. Physical realizability here means that the QSDEs should rep-

resent the dynamics of some meaningful physical system. For example, it could be that they represent the dynamics of the position and momentum operators of an optical cavity, a well-known device in quantum optics. Physical realizability is a very important issue from an engineering point of view since only implementable quantum systems, built from real physical devices, are of interest in real-life applications.

In the classical (non-quantum) setting, the question of realizability of deterministic and stochastic linear (time invariant) systems has been extensively studied, for example within the context of the theory of electrical networks, and is well understood. In principle, once all the coefficients are known, classical systems can be regarded as always being realizable, at least approximately, via a network of electrical, electronic and/or mechanical devices. For quantum stochastic systems, however, there are additional constraints that must be satisfied by the QSDEs to be physically meaningful, constraints which are not required of classical stochastic systems represented by a system of linear stochastic differential equations on a classical probability space. Among these constraints, physical systems are characterized by the preservation of canonical commutation relations (CCR) among certain canonical operators. The second part of the thesis introduces a formal notion of physical realizability for quantum linear stochastic systems, as well as deriving explicit necessary and sufficient conditions for preservation of the CCR and physical realizability. These conditions are relevant for extending the controller synthesis techniques of modern linear control theory, such as the LQG and H^∞ synthesis techniques, to the setting of quantum linear systems. The realization ideas are applied to a quantum H^∞ synthesis framework to show that controllers obtained from this synthesis can always be made to be physically realizable by appropriately adding a number of quantum noise channels. Moreover, the controllers can be freely chosen to be fully quantum (no classical components), fully classical (no quantum components), or a mixture of quantum-classical components.

Contents

Chapter 1

Introduction

Disturbances of a *random* or *stochastic* nature appear naturally in engineering practice and real-life applications. Noises in electrical and other devices, the erratic movement of prices in the stock market, turbulence that one experiences when travelling on an aircraft, are some examples of random disturbances which can be observed in the real-world. When these random effects are significant and cannot be neglected, it is more appropriate to consider the system of interest, be it a device, the stock market or an aircraft, as being stochastic, as opposed to being deterministic. Human endeavor to understand and to be able to quantitatively describe random processes in nature, so that its effects may be controlled to some degree or used to our advantage, has led to the development of the mathematical theory of stochastic processes which in turn provides the foundation for stochastic systems theory and stochastic control. For some systems, randomness is an indispensable part of its description. One case in particular are quantum mechanical systems which behave according to the laws of quantum mechanics. This is because randomness is an inherent feature built into the foundation of quantum mechanics, and quantum phenomena are most appropriately described using a probabilistic language.

This thesis considers two topics in the broad area of linear stochastic systems. One topic lies in the domain of traditional stochastic systems on a classical probability space, while the second topic is in the domain of quantum stochastic systems on a quantum probability space, an emerging area of research in engineering which is gaining more importance. The first topic is the construction of approximate finite dimensional linear time invariant models for classical stationary stochastic signals with a spectral density which is both non-coercive and non-rational. The second topic is physical realizability of linear quantum systems represented by

linear quantum stochastic differential equations. This chapter serves to give an overview of both topics and a summary of the contributions of the thesis. We now start with an overview of the first topic.

1.1 Approximate finite dimensional models for stochastic signals

Wide sense stationary (WSS) processes form an important class of stochastic processes. *Stochastic signals* which are system inputs or outputs are modelled as sample functions of *stochastic processes*, but in engineering contexts as here, the terms are loosely used *interchangeably*. Wide sense stationary stochastic processes are characterized as stochastic processes with time invariant first and second order statistical properties, that is, the mean and covariance respectively. Typically one sets the mean to be zero by adding an appropriate constant so that a WSS process may be assumed to be completely described by its covariance function. An equivalent complete description can be given in terms of the so-called spectral distribution of the process, which in most cases of practical interest is simply the Fourier transform of the covariance function. If the spectral distribution is absolutely continuous then its derivative is called the *spectral density* or the *spectrum* of the process. A striking result from the theory of WSS processes is that such processes having a spectral density (i.e., the spectral distribution is absolutely continuous) can always be modelled as the output of a linear time invariant (LTI) system (referred to as a "shaping filter") driven by a white noise input [1]. This provides a universal model of WSS processes as well as a practical way of generating them, i.e., by identifying an appropriate shaping filter.

The operation of obtaining an LTI model from a spectral density is known in the literature as *spectral factorization* and plays an important role in both deterministic and stochastic linear systems theory. More generally, spectral factorization and the results presented in this thesis are not only of interest for modelling of stochastic signals and systems, but also for the optimal control of some classes of deterministic infinite dimensional/distributed parameter systems. For such systems, spectral factorization has been developed in the literature as one technique for synthesizing optimal controllers [2, 3, 4]. A more detailed discussion of WSS processes and spectral factorization will be given in Chapter 2.

In the thesis we will mainly be concerned with discrete time wide-sense stationary processes with a spectral density, defined on the unit circle, which is

non-coercive and non-rational. Non-coercive means that the spectral density has one or more zeros on the unit circle, while non-rationality implies that the underlying system generating such a signal has infinite dimension (also known in the literature as distributed parameter systems). Nonetheless, our results can also be applied to continuous time processes by application of a standard bilinear transformation which maps the imaginary axis onto the unit circle.

An attribute of the class of spectral densities mentioned above, which we shall shortly elaborate upon, is that, in general, their spectral factors are difficult to compute numerically. Given such a scalar spectral density W, a particular rational spectral density W_n, known as a *maximum entropy* spectrum, can be constructed whose first n terms of the covariance sequence matches those of W [5, 6]. Matching of partial covariance sequences is a natural approach since, as mentioned briefly in the previous paragraph, every spectral density is associated to a unique covariance sequence, and both give a complete and equivalent characterization of some purely non-deterministic wide sense stationary stochastic process [7, 1]. H_n, the canonical spectral factor of W_n, can be constructed recursively via the Szegö-Levinson algorithm, see, e.g., [8, 7], and it is well-known that as $n \uparrow \infty$, H_n converges to H, the canonical spectral factor of W, in $\mathcal{H}^2$, the Hardy space of functions square-integrable on the unit circle and having vanishing negative Fourier coefficients. Therefore, also W_n converges to W in $\mathcal{L}^1$, the space of integrable functions on the unit circle. However, H_n obtained in this way is an all-pole transfer function. Since H_n has no zeros, it has long been observed that $H_1, H_2, \ldots$ converges slowly to H when the latter has zeros on or close to the unit circle, some examples can be found in, e.g., [9, Section IV], [5, pp. 214-217] and [6, Section 6]. Intuitively, this is due to the inability of W_n to reproduce valleys of W (i.e., points for which W has a small value) for small or medium n. Similar slow convergence is also true when the matrix generalization of the Szegö-Levinson algoritm (see [10, 11]) is applied to a matrix-valued spectrum W with transmission zeros on or close to the unit circle [10]. More formally, it was shown in [12, 9] that if W is scalar and rational, then its zeros that are close to or on the unit circle decreases the rate of decay of the Schur parameters of W: convergence rate decreases as a zero approaches the unit circle (see also [13]). In particular, when W has roots on the unit circle, the rate of convergence is no longer geometric. Consequently, good approximations can only be achieved for large n.

The limitation of the maximum entropy spectra motivated research into development of rational spectral densities which 1) matches the first few terms of the

covariance sequence of W and 2) has zeros at specified locations on the complex plane. A new theory of rational covariance extension with degree constraint was later developed which makes possible the construction of such spectra. Indeed, it has been demonstrated in [5, 6] that a finite dimensional spectrum constructed via the new theory, by suitably placing zeros on the unit disc, is a better estimate than the corresponding maximum entropy estimate, in the sense that it is able to better capture features of the true spectrum.

The first part of the thesis (Chapters 2-6) establishes some new results for degree constrained rational covariance extension and interpolation, a general problem with ties to many applications in systems and control [14], and then develops a new approach to spectral factorization of a spectral density W, based on constructing a sequence of rational approximations $\{W_n\}$ with freely specified zeros and which match partial covariance sequences of W. The construction is achieved by taking advantage of a recent result on continuity of the spectral factorization mapping given in [15] and the theory of degree constrained rational covariance extensions [12, 9, 16, 5, 6, 17, 18]. We derive theoretical results on convergence of this scheme for continuous W, and conditions on the zeros for convergence to be achieved, under further mild assumptions on W. Convergence of covariance matching approximations with freely specified zeros have not been studied in the literature; convergence results have only been established for the case where W_n has no zeros, i.e., the maximum entropy method/Szegö-Levinson algorithm. In particular, our results weaken some conditions previously derived by Anderson [19], Caines and Baykal-Gürsoy [20], and Mari et al [13] for convergence of $\{H_n\}$ to H in $\mathcal{H}^2$ and $\mathcal{H}^\infty$, respectively, where $\mathcal{H}^\infty$ denotes the space of functions which are analytic and bounded on the open unit disc. Then in the penultimate section of Chapter 6, several numerical examples are given that demonstrate the advantage of the proposed approach over the popular maximum entropy method: lower degree approximations with lower approximation error (to be defined in a certain sense).

From an applications side, non-coercive and non-rational spectral densities are of particular interest in optics, astronomy and flight research. As one example, spectral densities of this type appear when measurements are taken of signals that have traversed through the Earth's turbulent atmosphere, such as light from a distant star captured by astronomical telescopes on the ground. In the literature, analytical derivations based on the so-called Taylor frozen hypothesis [21] show that signals captured by wavefront sensors on telescopes are continuous time signals with non-coercive (for continuous time this means that the spectral density

has zeros on the imaginary axis) and non-rational temporal spectral densities, see [22] and the references cited therein. These properties agree with empirical observations reported in, for example, [23, 24, 25]. Knowledge of the temporal spectra is of interest for estimating and improving the performance of various high resolution imaging systems which compensate for abberation of images received on a telescope due to atmospheric turbulence [22, 25]. One such image enhancing system which has gained a significant amount of attention in recent years are adaptive optics control systems [21, 26, 27]. In adaptive optics, deformable mechanical mirrors are shaped according to some control algorithm to correct the phase of the wavefront of the incident light, which has been distorted from its original planar profile by atmospheric turbulence, before an image is formed on the telescope. As another relevant example, the spectral densities of the components of wind turbulence velocity are also modelled to be of the non-coercive and non-rational type [28, 29]. Wind turbulence models are important for computer-based simulation of flight conditions in the design of aircrafts and its sub-systems, including the flight controller. Typical approximate models that are used in current applications are low order ones which give a good fit to the spectral density only in a limited frequency range [29, 26, 27], whereas the method proposed here allows one to obtain higher degree models which give a better fit across a wider range. The choice of the appropriate higher degree model would then be application specific, depending on the particular constraints present, such computational constraints.

In summary, the ideas and results presented in the first part of the thesis may be useful in applications in which spectral factorization plays a prominent role such as in computation of approximate solutions of algebraic Riccati equations (ARE's) in optimal control of linear systems, or in which signals with non-rational power spectra is a central theme (e.g., control of aircraft subject to windgust and adaptive optics as discussed above, and also in laser scintillation [30]). They may also prove to be useful in spectral estimation and system identification research.

1.2 Physical realizability of quantum linear stochastic systems

The second topic addressed is the issue of physical realizability of a given system represented by linear quantum stochastic differential equations (QSDEs) on a *quantum probability space*.

Physical realizability here means that the QSDE represents the dynamics of some meaningful physical system. For example, it could be that the QSDE represent the dynamics of the position and momentum operators of an optical cavity, a well-known device in quantum optics. Physical realizability is a very important issue from an engineering point of view since only implementable quantum systems, built from real physical devices, are of use in real-life applications.

The realization question for classical (i.e., non-quantum) linear systems represented by a system of ordinary linear differential equations is a central one in engineering and has been extensively studied in the literature. In electrical engineering, this question is addressed in the context of synthesis of electrical networks, at first using classical complex function theory and later with modern state-space methods [31]. However, the results for electrical networks can, by analogy, be adapted to other kinds of networks. For a comprehensive treatment of the state-space approach for synthesis of electrical networks, see [31]. According to the theory, we may, in principle, always regard linear deterministic or stochastic linear (time invariant) systems as being realizable via a network of electrical, electronic and, possibly, mechanical devices. This is not the case for quantum linear stochastic systems. There are additional constraints, not present in the non-quantum context, that must be satisfied by the QSDEs for them to be physically meaningful. One such constraint is the requirement that certain canonical commutation relations (CCR) between canonical conjugate variables of the QSDEs should be satisfied at all times. The purpose of the second part of this thesis is to introduce a formal notion of physical realizability for quantum linear stochastic systems and to derive necessary and sufficient conditions for preservation of the CCR and physical realizability. Our setting includes the interesting and important case of systems which may have both classical and quantum degrees of freedom, and includes linear models which are of interest in quantum optics.

With the recent advances, both theoretical and experimental, in the control of objects at the nano scale, such as control of an atom by using a modulated laser, and potential benefits that may be offered by quantum information and signal processing system over their classical counterparts, quantum control has become a significant topic which has recently attracted more attention. Due to the important role that linear systems has played in the development of classical control theory, a similar theory for quantum linear systems could be beneficial for better understanding of quantum control theory. The results of the second part of this thesis takes a step in that direction by building a connection between

abstract physical concepts with realizability issues which are important from an engineering perspective. In particular, the ideas developed are applied in the context of an H^∞ synthesis framework due to James and Petersen [32] to show that the controllers which result from this synthesis can always be made to be physically realizable.

1.3 Summary of contributions of the thesis

The original contributions of this thesis are as follows:

- Some new results on degree constrained rational interpolation are established for the case where the parametrizing pseudopolynomial has spectral zeros on the unit circle. This completes the analysis for the dual side of the primal-dual convex optimization approach pioneered by Byrnes et al [5, 6]. The results include:

    1. A necessary and sufficient condition for a degree constrained rational interpolant to be bounded on the unit disc (Chapter 3).
    2. A characterization of the denominator polynomial of all degree constrained rational interpolants, including unbounded ones having a pole of the unit circle (Chapter 4). This leads to several corollaries, including one which establishes a homeomorphism between the numerator and denominator polynomials of an interpolant and its pair of partial covariance sequence and parametrizing pseudopolynomial.

- It is shown that a numerically stable homotopy continuation method, originally proposed by Enqvist for computing degree constrained rational covariance extensions which are bounded and strictly positive real, is actually applicable for all interpolants including those that are unbounded or non-strictly positive real (Chapter 5). This establishes that the algorithm provides a complete computational method, and it has the potential to be extended to interpolation problems beyond rational interpolation.

- A new approach to spectral factorization of a class of spectral densities is introduced (Chapter 6) based on a sequential continuity property of the spectral factorization mapping and the ideas and techniques developed in Chapters 3 and 4. A new spectral factorization algorithm is also introduced

and analyzed. Several numerical simulations are provided to indicate the performance of the algorithm.

- A notion of physical realizability for a class of quantum linear stochastic systems is developed, and characterizations of physical realizability are derived (Chapter 7).

- In the context of H^∞ synthesis, it is shown that physically realizable robust controllers (Chapter 8) always exist. Moreover, it is shown that the controllers may be freely specified to consist of either purely quantum degrees of freedom, purely classical degrees of freedom, or a mixture of both quantum and classical degrees of freedom.

1.4 Organization of the thesis

The organization of this thesis is as follows. It is divided into two main parts, with Part I dealing with the first topic (chapters 2 to 6) and Part II (Chapters 7 and 8) dealing with the second topic.

In Chapter 2, we review some concepts from wide sense stationary processes to motivate spectral factorization. In particular, we show the relationship of spectral factorization to the prediction theory of stationary processes and to innovation and modelling filters. We also discuss the so-called Szegö-Levinson algorithm for solving the prediction problem and its generalization, the (generalized) Schur algorithm which is the basis for the Darlington synthesis procedure in circuit theory. We point out some limitations of the Schur algorithm for spectral factorization. In Chapter 3 we give an exposition on the rational covariance extension problem with degree constraint (RCEP) and its bounded solutions. Chapter 4 continues the development in Chapter 3 by analyzing all solutions of the RCEP, including unbounded ones, and showing a homeomorphic correspondence between pairs of partial covariance sequence and non-negative pseudopolynomials data with positive real rational functions of a bounded degree. In Chapter 5, results from Chapter 4 are exploited to establish that a numerical homotopy continuation algorithm originally due to Enqvist is in fact also applicable for computing degree constrained rational interpolants corresponding to parametrizing pseudopolynomials with spectral zeros on the unit circle, a case which has not previously been studied for this algorithm. Then in Chapter 6 we introduce a new framework for spectral factorization for a certain class of spectral densities based on results from

[15] and ideas developed in Chapters 3 and 4. The class allows for the possibility of non-coercive spectral densities (i.e., those having zeros on the unit circle) which are known to be particularly difficult to factorize for many spectral factorization algorithms. A new spectral factorization algorithm is also introduced and convergence results provided. The effectiveness of the algorithm are demonstrated in a number of numerical examples. In two examples, we apply the algorithm to compute approximate spectral factors of the "physically derived" non-coercive and non-rational von Karman and Kolmogorov spectral densities which arise in the study of atmospheric turbulence.

In Chapter 7, we develop a notion of physical realizability for quantum linear stochastic systems which are of interest in quantum optics, and derive characterizations of physically realizable quantum linear stochastic systems. Then in Chapter 8, we describe an H^∞ controller synthesis framework for quantum linear stochastic systems which gives a partial description of a controller and show that such a partial description may always be completed such that the resulting controller is physically realizable in the sense of Chapter 7. Synthesis examples in the context of quantum optics are given in which different types of controllers are realized, including a fully quantum controller (with no classical component), a fully classical controller (with no quantum component), and a mixed quantum-classical controller.

Part I

Topics in Classical Linear Stochastic Systems

List of Notation and Terminology for Part I

Notation

$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
$\mathbb{D}$	The unit disc $\{z \in \mathbb{C} \mid \lvert z\rvert < 1\}$
$\mathbb{T}$	The unit circle $\{z \in \mathbb{C} \mid \lvert z\rvert = 1\}$
$\operatorname{col}(a_1, \ldots, a_n)$	$[a_1 \ \ldots \ a_n]^T$
$\overline{A}$	Depending on the context, denotes either the closure or completion of A, or the elementwise complex conjugation of a complex matrix A
∂A	The boundary of a topological set A
A^*	The conjugate transpose of a complex matrix A
$\Re\{A\}$	$\Re\{A\} = A + A^*$, the hermitian part of a complex matrix A
$\mathbb{Z}$	The set of integers $\ldots, -3, -2, -1, 0, 1, 2, 3, \ldots$
$\mathbb{N}$	The set of natural numbers $1, 2, 3, \ldots$
μ	The Lebesque measure on $\mathbb{T}$
$\lVert A\rVert_p$	The (Schatten) p-norm a matrix $A \in \mathbb{C}^{m\times n}$ defined as: $$\lVert A\rVert_p = \begin{cases} \left(\operatorname{Tr}\left\{(A^*A)^{p/2}\right\}\right)^{\frac{1}{p}} & \text{if } 1 \le p < \infty, \\ \sup_{v\in\mathbb{C}^n, \lVert v\rVert\le 1} \lVert Av\rVert & \text{if } p = \infty. \end{cases}$$

$\mathcal{L}^p_{m\times n}, 1 \le p \le \infty$ The space of measurable functions mapping from $\mathbb{T}$ to $\mathbb{C}^{m\times n}$ with a finite $\|\cdot\|_p$ norm defined by:

$$\|f\|_p = \begin{cases} \left(\frac{1}{2\pi}\int_{\mathbb{T}} \|f(z)\|_p^p d\mu\right)^{\frac{1}{p}} & \text{if } 1 \le p < \infty \\ \operatorname{ess\,sup}_{z\in\mathbb{T}} \ \|f(z)\|_\infty & \text{if } p = \infty \end{cases}$$

If $n = 1$, then $\mathcal{L}^p_{m\times n}$ is denoted simply as $\mathcal{L}^p_m$

$\mathcal{H}^p_{m\times n}$, $1 \le p \le \infty$ The subspace of functions in $\mathcal{L}^p_{m\times n}$ having an analytic continuation from $\mathbb{T}$ to $\mathbb{D}$. If $n = 1$, $\mathcal{H}^p_{m\times n}$ is denoted simply as $\mathcal{H}^p_m$

H_* The *parahermitian conjugate* of a $\mathbb{C}^{m\times n}$-valued complex function H defined by $H_*(z) = H(z^{*-1})^*$

Terminology

Outer function A function $H \in \mathcal{H}^2_{n\times n}$, $n \in \mathbb{N}$, such that the set $\{H\rho \mid \rho$ is a $\mathbb{C}^n$-valued polynomial in $\mathbb{C}\}$ is dense in $\mathcal{H}^2_{n\times n}$

Pseudopolynomial A complex function of the form

$$f(z) = a_0 + \sum_{k=1}^{n} \left(a_k^* z^{-k} + a_k z^k\right),$$

where $0 \le n < \infty$, $a_n \ne 0$ and $(a_0, a_1, \ldots, a_n) \in \mathbb{R} \times \mathbb{C}^n$. n is the *order* or *degree* of the pseudopolynomial f (the order is zero if f is a constant function). A matrix-valued pseudopolynomial may also be defined by letting $(a_0, a_1, \ldots, a_n) \in \mathbb{R}^{l\times l} \times \mathbb{C}^{l\times l} \times \ldots \times \mathbb{C}^{l\times l}$ for some integer $l > 1$

Spectral density A function W in $\mathcal{L}^1_{n\times n}$, for some $n \in \mathbb{N}$, satisfying $W(e^{i\theta}) = W(e^{i\theta})^*$, $W(z) \ge 0$ for almost all $z \in \mathbb{T}$, and

$$\int_{\mathbb{T}} |\log\det W(z)| \mu(dz) < \infty$$

Spectrum Another term for spectral density

Chapter 2

Wide-Sense Stationary Processes and Spectral Factorization

2.1 Introduction

The purpose of this chapter is to give a brief review of the theory of wide-sense stationary (which we shall again abbreviate as WSS as in Chapter 1) processes. Of particular importance is the relationship between prediction theory of WSS processes, spectral factorization and the so-called Darlington synthesis procedure of circuit theory. We shall explain why some of the more popular and general methods for spectral factorization are inadequate for "harder" spectral densities which have zeros close to or on the unit circle. For a comprehensive treatment of the topics of this chapter we refer the reader to texts such as [33, 1, 34, 7] and the papers [8, 35, 36, 37, 9, 38].

2.2 Second order and wide-sense stationary processes

By second order processes, we actually mean discrete time second order stochastic processes. In the literature, second order stochastic processes are studied in both continuous and discrete time, but our main interest will be the discrete time setting. Although the continuous time theory is essential for theoretical study of real-life processes, in applications filtering and control algorithms are typically implemented on digital devices, such as microprocessors and high-speed digital signal processors (DSPs), which inherently operate in discrete-time.

Let $\mathbb{R}$, $\mathbb{C}$, $\mathbb{D}$ and $\mathbb{T}$ denote the real numbers, complex numbers, the unit disc $\{z \in \mathbb{C} \mid |z| < 1\}$ and the unit circle $\{z \in \mathbb{C} \mid |z| = 1\}$, respectively. A second order process is defined as a sequence of $\mathbb{C}^n$-valued random variables $\{X_k\}_{k\in\mathbb{Z}}$ ($\mathbb{Z}$ denotes the set of all integers) defined on some probability triplet $(\Omega, \mathcal{F}, P)$, i.e., Ω is the set of events, $\mathcal{F}$ is a σ-algebra of subsets of Ω and P is a probability measure on $\mathcal{F}$, satisfying $\mathrm{Tr}\{\mathrm{E}(X_k - \mathrm{E}X_k)(X_k - \mathrm{E}X_k)^*\} < \infty$ for all $k \in \mathbb{Z}$ (we assume that elements of $\mathbb{C}^n$ are represented as column vectors), where $\mathrm{E}Y$ denotes the expectation of a random variable Y with respect to P and * denotes conjugate (or Hermitian) transposition of a matrix. Define $m(k) = \mathrm{E}X_k$ and $R(k,l) = \mathrm{E}(X_k - m(k))(X_l - m(l))^*$. Then $R(k,l)$ is called the *covariance function* of the process $\{X_k\}_{k\in\mathbb{Z}}$. It is easy to see that $R(l,k) = R(k,l)^*$, and $R(k,l)$ possesses the non-negative definite property in the sense that:

$$\sum_{k=1}^{m}\sum_{l=1}^{m} a_k^* R(i_k, i_l) a_l \geq 0,$$

for any positive integer m and for any arbitrary collection $i_1, \ldots, i_m$ of integers and any arbitrary collection of complex numbers $a_1, \ldots, a_m$. If $R(k,l) \equiv 0$ then we say that X_k and X_l are (mutually) *orthogonal* or *uncorrelated*. We write this as $X_k \perp X_l$.

WSS processes form a special, yet important, class of second order processes. This special class has the additional property that $\mathrm{E}X_k = c \ \forall k \in \mathbb{Z}$, c being a complex constant, and $R(k,l) = r(k-l)$ for some bounded matrix-valued function r defined on $\mathbb{Z}$ satisfying $r(-k) = r(k)^*$. For WSS processes we say that $r(0), r(1), \ldots$ is the *covariance sequence* of the process. It is straightforward to check that $|\mathrm{Tr}\{r(k)\}| \leq \mathrm{Tr}\{r(0)\}$ for all $k \in \mathbb{Z}$, and it is also non-negative definite in the sense that:

$$\sum_{k=1}^{m}\sum_{l=1}^{m} a_k^* r(i_k - i_l) a_l \geq 0,$$

for any positive integer m and for any arbitrary collection $i_1, \ldots, i_m$ of integers and any arbitrary collection of complex numbers $a_1, a_2, \ldots, a_m$. Moreover, there will be no loss in generality in assuming that $c = 0$, so for convenience we shall make this assumption for the remaining of this chapter.

The special properties of WSS processes allows the development of a rich mathematical theory for these processes. We shall now give a review of this theory starting with a discussion of the Hilbert space structure of WSS processes.

2.3 Hilbert space structure of WSS processes

For the ease of discussion, we shall assume for the time being that $\{X_k\}_{k\in\mathbb{Z}}$ is a scalar WSS process. First, we may interpret $\mathrm{E}X_kX_l^*$ as a complex inner-product between X_k and X_l. The inner product is also well defined for any finite linear combination of the X_k's. Denoting the inner product as $\langle\cdot,\cdot\rangle_X$, we have:

$$\begin{aligned}\left\langle\sum_{k=1}^m \alpha_k X_{i_k}, \sum_{l=1}^m \beta_l X_{i_l}\right\rangle_X &= \mathrm{E}\left(\sum_{k=1}^m \alpha_k X_{i_k}\left(\sum_{l=1}^m \beta_l X_{i_l}\right)^*\right),\\ &= \sum_{k=1}^m\sum_{l=1}^m \alpha_k\beta_l^*\mathrm{E}X_{i_k}X_{i_l}^*,\\ &= \sum_{k=1}^m\sum_{l=1}^m \alpha_k\beta_l^*\langle X_{i_k}, X_{i_l}\rangle_X.\end{aligned}$$

Let $h(X)$ denote the span of $\{X_k\}_{k\in\mathbb{Z}}$ over $\mathbb{C}$, i.e., the set of all possible finite linear combinations of the X_k's over the field $\mathbb{C}$. Then the inner product $\langle\cdot,\cdot\rangle_X$ induces a norm $\|\cdot\|_X$ on $h(X)$ defined by $\|Y\|_X = \langle Y, Y\rangle_X^{\frac{1}{2}}$. By continuity of the inner product, it can be linearly extended from $h(X)$ to $H(X) = \overline{h(X)}$, where $\overline{h(X)}$ denotes the completion of $h(X)$ with respect to the norm $\|\cdot\|_X$. Then $H(X)$ is a complete normed space with inner product $\langle\cdot,\cdot\rangle_X$ [1], i.e., $H(X)$ is a Hilbert space. This allows one to conveniently analyze WSS processes within a purely geometric framework. In particular, optimization becomes simple since the projection of an element Y onto any closed subspace of $H(X)$ is uniquely defined. The Hilbert space structure of WSS processes led H. Wold to formulate the Wold decomposition. This decomposition is of fundamental importance in time series analysis and will be discussed in the next section.

For vector-valued (or even matrix-valued) processes only some simple adaptations are necessary. Suppose that X_k is a $\mathbb{C}^n$-valued random variable for each k and let us write $X_k = \mathrm{col}(X_{k,1},\ldots,X_{k,n})$, where $X_{k,j}$ is a scalar random variable for $1 \le j \le n$. Since $\{X_k\}_{k\in\mathbb{Z}}$ is a stationary process, it follows by the same arguments as in the scalar case that $H(X) = \overline{\mathrm{span}\{\cup_{k\in\mathbb{Z}}\{X_{k,1},\ldots,X_{k,n}\}\}}$ is also a Hilbert space. Therefore, exactly as in the scalar case, we may apply operations such as taking the unique projection of any element of $Y \in H(X)$ onto some closed subspace of $H(X)$, e.g., taking the projection of $X_{1,1}$ onto $\mathrm{span}\{X_{1,n_1}, X_{2,n_2},\ldots,X_{k,n_k}\}$, where the n_k's may be assigned arbitrary values in $\{1,\ldots,n\}$.

2.4 The Wold decomposition, prediction of stationary processes and spectral factorization

Let $\{X_k\}_{k\in\mathbb{Z}}$ be a WSS process. We once again assume that the process is scalar. However, note that most of the ideas and results to follow carry over to vector-valued processes (see, for example, the classical papers [39, 40, 41, 42]).

Define $H_k(X) = \overline{\text{span}\{X_l\}_{l\leq k}}$ and $H_{-\infty}(X) = \cap_{k\in\mathbb{Z}} H_k(X)$. Then we say that $\{X_k\}_{k\in\mathbb{Z}}$ is *deterministic* if $X_0 \in H_{-1}$. Otherwise, the process is non-deterministic. The Wold decomposition theorem, due to H. Wold, says that any non-deterministic stationary process can be decomposed into the sum of two mutually uncorrelated deterministic and *purely non-deterministic* (PND), to be defined shortly, processes. To be precise:

Theorem 2.4.1 *Any non-deterministic process $\{X_k\}_{k\in\mathbb{Z}}$ has the following decomposition:*

$$X_k = U_k + V_k,$$

where $V_k \in H_{-\infty}(X)\ \forall k$ and $\{V_k\}_{k\in\mathbb{Z}}$ is a deterministic process, $U_k \perp V_l\ \forall k, l$ and

$$U_k = \sum_{l=-\infty}^{k} b_{k-l} e_l, \tag{2.1}$$

where $e_l \in H_l(X)$ and $\|e_l\|_X = \sigma > 0\ \forall l \in \mathbb{Z}$, $e_i \perp e_j\ \forall i \neq j$, $b_0 = 1$, $\sum_{l=1}^{\infty} |b_l|^2 < \infty$, and $\langle e_k, X_{k+l}\rangle_X = \sigma^2 b_l\ \forall l \geq 0$.

In the literature, the random variables $\ldots, e_{-1}, e_0, e_1, \ldots$ as defined in the theorem are often referred to as *innovations*, and the process $\{e_k\}_{k\in\mathbb{Z}}$ as the *innovation process*. Moreover, any process U which can be decomposed in the manner of the right hand side of (2.1) is said to be a purely non-deterministic process.

The significance of the Wold decomposition theorem was recognized by A. N. Kolmogorov, who subsequently developed a comprehensive theory of prediction of (scalar) stationary processes. The central theme of this theory is that of best *linear* predictors/estimators, and the full generality of probability theory may be dispensed of by working exclusively with Hilbert space theory. However, unless the process follows a jointly Gaussian distribution, the predictors obtained by this theory will not be optimal. This is because, in general, there will exist a non-linear predictor that outperforms the best linear one.

Kolmogorov observed that since $e_i \perp e_j \ \forall i \neq j$ then the best linear estimator of X_k in $H_{k-1}(X)$ given $\{X_l\}_{l<k}$ must necessarily be $X_k - e_k$ for which the estimation error is then given by e_k. In his work on prediction theory Kolmogorov showed the following:

Theorem 2.4.2 *The following are equivalent:*

1. $\{X_k\}_{k\in\mathbb{Z}}$ *is WSS.*

2. *There exists a unitary operator* U *on* $H(X)$ *such that* $X_k = U^k X_0 \ \forall k \in \mathbb{Z}$.

Note that a unitary operator U is an operator satisfying $U^* = U^{-1}$, where U^* denotes the adjoint operator of U satisfying the relation $\langle Ux, y\rangle_X = \langle x, U^* y\rangle_X$ $\forall x, y \in H(X)$. It is then shown that there exists a *spectral measure* E (see [7]) such that the unitary operator U can be conveniently expressed as:

$$U = \int_{-\pi}^{\pi} e^{i\lambda} dE(\lambda),$$

where the above integral is defined in the sense of the identity:

$$\langle Ux, y\rangle_X = \int_{-\pi}^{\pi} e^{i\lambda} \langle dE(\lambda)x, y\rangle_X,$$

$\forall x, y \in H(X)$. Then $U^k x = \int_{-\pi}^{\pi} e^{ik\lambda} dE(\lambda)x$ for any $x \in H(X)$ and the *spectral representation* of WSS processes can be established:

Theorem 2.4.3 *For a WSS process* $\{X_k\}_{k\in\mathbb{Z}}$ *there exists a process* $\{Z(\lambda); -\pi < \lambda \leq \pi\}$ *with orthogonal increments (i.e.,* $Z(\lambda_2) - Z(\lambda_1) \perp Z(\lambda_4) - Z(\lambda_3)$ *for any* $-\pi < \lambda_1 < \lambda_2 < \lambda_3 < \lambda_4 \leq \pi$*) such that:*

$$X_k = \int_{-\pi}^{\pi} e^{ik\lambda} dZ(\lambda) \quad \forall k \in \mathbb{Z}, \tag{2.2}$$

where the integral above is to be interpreted as a stochastic integral.

For details on the stochastic integral, see, e.g., [7, 1]. The integral representation of X_k in terms of the Z process in the theorem is referred to as the *spectral representation* of X_k. From the representation we may show that:

$$\begin{aligned} \langle X_k, X_l\rangle_X &= \int_{-\pi}^{\pi} e^{i(k-l)\lambda} \|dZ(\lambda)\|_X^2, \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{ik\lambda} e^{-il\lambda} dF_X(\lambda), \end{aligned} \tag{2.3}$$

where F_X is a measure called the *spectral distribution* of $\{X_k\}_{k\in\mathbb{Z}}$. F_X has the property that $F_X((-\pi,\lambda])$ is a non-decreasing and *right continuous* function of λ on $(-\pi,\pi]$ with $F_X(\pi) = r(0)$. It also has the decomposition $F_X = F_U + F_V$, where F_U and F_V are absolutely continuous and singular with respect to the Lebesque measure on $(-\pi,\pi]$, respectively. Therefore, $dF_U(\lambda) = f_X(\lambda)d\lambda$ for some function f_X which in integrable on $(-\pi,\pi]$. Clearly, $f_X(\lambda)$ must be non-negative definite on $(-\pi,\pi]$. If the process is PND then $F_V \equiv 0$ and f_X is referred to as the *spectral density* of $\{X_k\}_{k\in\mathbb{Z}}$.

From (2.3) we have that:

$$\left\langle \sum_{k=1}^{m} a_{r_k} X_{r_k}, \sum_{l=1}^{m} b_{s_l} X_{s_l} \right\rangle_X = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left(\sum_{k=1}^{m} a_{r_k} e^{ir_k\lambda} \right) \left(\sum_{l=1}^{m} b_{s_l} e^{is_l\lambda} \right)^* dF_X(\lambda), \tag{2.4}$$

holds true for all finite m and for all possible sequences $r_1,\ldots,r_m$ and $s_1,\ldots,s_m$ taking values in $\mathbb{Z}$ and for any sequence $a_1,\ldots,a_m$ and $b_1,\ldots,b_m$ taking values in $\mathbb{C}$. Note that the right hand side of (2.4) defines an inner product on $p(F) = \text{span}\{e^{ik\lambda}\}_{k\in\mathbb{Z}}$. The linear map K defined via $K : x_k \mapsto e^{ik\lambda}$ defines an isomorphism form $h(X)$ to $p(F)$ which preserves the inner product. In other words, K is an isometry. Then by continuity of the inner product, K can be extended to an isometry from $H(X)$ to $P(F) = \overline{p(F)}$. The isomorphism K is known as the *Kolmogorov isomorphism*. It allows one to convert least squares minimization problems in $H(X)$ to least squares approximation problems in $P(F)$. There are certain analytical properties of $P(F)$ which makes the latter convenient for analyzing such problems and it is a key feature of Kolmogorov's work on prediction theory. The main results are as follows:

Theorem 2.4.4 *Let $\{X_k\}_{k\in\mathbb{Z}}$ be WSS. Then:*

1. *$\{X_k\}_{k\in\mathbb{Z}}$ is a non-deterministic process if and only if*

$$\int_{-\pi}^{\pi} \log f_X(\lambda)d\lambda > -\infty.$$

   If, in addition, F_V is the zero measure, then it is PND.

2. *For a non-deterministic process, the one step ahead prediction error variance $\sigma^2 = \|e_l\|_X$ is given by:*

$$\sigma^2 = \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \log f_X(\lambda)d\lambda \right)$$

3. *If* $c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{ik\lambda} \log f_X(\lambda) d\lambda$ *for* $k \in \mathbb{Z}$, *then* $b_1, b_2, \ldots$ *in Theorem 2.4.1 satisfies:*

$$\sigma(1 + b_1 z + b_2 z^2 + \ldots) = \psi(z) = \exp\left\{ \frac{c_0}{2} + \sum_{j=1}^{\infty} c_j z^j \right\}, |z| < 1,$$

where $|\psi(e^{i\lambda})|^2 = f_X(\lambda)$.

Assuming that the process is PND, the function ψ defined in the last theorem is called a *spectral factor* of f_X. It can be shown that it must be an *outer function* (i.e., it is analytic inside the unit disc and has no zeros there) and is unique up to multiplication by a complex number of modulus one. The operation of finding such a function ψ for a given f_X is referred to as *spectral factorization*. The last theorem gives us an important connection between spectral factorization and the prediction theory of stochastic processes. More details on this relationship are given in the next section.

2.5 Finite prediction and the Szegö-Levinson algorithm

The work of Kolmogorov does not give formulas for the one step-ahead predictor $\hat{X}_{k+1|\infty}$ of X_{k+1} in terms of *all* present and past observed values $X_k, X_{k-1}, \ldots$, but in terms of the unobservable innovations $e_k, e_{k-1}, \ldots$. A method to approximate the predictor is to compute the best (linear) estimate of X_{k+1} given the *partial* past $X_k, X_{k-1}, \ldots, X_{k-n}$ (i.e., the projection of X_{k+1} onto $H_k(X) \backslash H_{k-n-1}(X)$) and then letting $n \to \infty$. Thus, one tries to solve a sequence of the following problem:

Problem 2.5.1 *Given a positive* $n \in \mathbb{Z}$, *find the unique vector* $(a_{n,0}, a_{n,1}, \ldots, a_{n,n}) \in \mathbb{C}^{n+1}$ *which minimizes* $\mathrm{E}|X_{k+1} - \sum_{l=0}^{n} a_{n,l} X_{k-l}|^2$.

By the Kolmogorov isomorphism the last problem is equivalent to the following:

Problem 2.5.2 *Given a positive* $n \in \mathbb{Z}$, *find the unique polynomial* $a_n(z) = a_{n,0} z^n + a_{n,1} z^{n-1} + \ldots + a_{n,n}$ *of degree at most* n *which minimizes the integral*

$$\int_{-\pi}^{\pi} |e^{i(k+1)\lambda} - e^{i(k-n)\lambda} a(e^{i\lambda})|^2 dF_X(\lambda) = \int_{-\pi}^{\pi} |e^{i(n+1)\lambda} - a_n(e^{i\lambda})|^2 dF_X(\lambda)$$

For the corresponding n, let $\hat{a}_n$ denote the solution to Problem 2.5.1 or Problem 2.5.2 (in the former problem $\hat{a}_n$ would be a vector while in the latter it would be a polynomial). N. Levinson showed that $\hat{a}_{n+1}$ in Problem 2.5.1 can be computed recursively and efficiently by updating $\hat{a}_n$ using new covariance data $r(n+1)$. However, many years before G. Szegö had already given a recursive solution to Problem 2 (hence by the isomorphism one automatically gets a recursion for the other problem) in his study of polynomials which are mutually orthogonal with respect to a distribution on the unit circle (known as Szegö's orthogonal polynomials). An exposition of orthogonal polynomials on the unit circle can be found in [33, 43]. theory of

Now, let ϕ_n denote the n-th Szegö orthogonal polynomial with respect to the distribution F_X and write it as $\phi_n = \sum_{k=0}^{n} \phi_{n,k} z^{n-k}$. Some properties of ϕ_n are as follows [33, 43]:

1. $\frac{\phi_n}{\phi_{0,n}}$ minimizes the integral $\frac{1}{2\pi}\int_{-\pi}^{\pi} |a(e^{i\lambda})|^2 dF_X(\lambda)$ over all monic polynomials a of degree n (i.e., the coefficient of z^n is 1). Moreover the minimum value of this integral is $\frac{1}{\phi_{0,n}^2}$.

2. ϕ_n has all zeros inside the unit circle $\mathbb{D}$.

3. $\lim_{n\to\infty} \|\phi_{n*} g - 1\|_2 = 0$, where g is a function in $\mathcal{H}^2$ satisfying $g(0) > 0$ and $|g(e^{i\lambda})|^2 = f_X(\lambda)$ for almost all $\lambda \in (-\pi, \pi]$.

Note that in the last property ϕ_{n*} denotes the *parahermitian conjugate* of ϕ_{n*} defined by $\phi_{n*}(z) = \phi_n(z^{*-1})^*$. The last property shows that one may consider $\frac{1}{\phi_*}$ as an approximate spectral factor of $f_X(\lambda)$. Now we shall briefly look at the relation between orthogonal polynomials and prediction theory of WSS processes.

It is well-known that the solution $\hat{a}_n$ to Problem 2.5.2 is given by:

$$\hat{a}_n(z) = -\sum_{k=0}^{n} u_{k+1}^* z^{n-k},$$

where the coefficients $u_1, u_2, \ldots, u_{n+1}$ are defined via the relation:

$$\begin{aligned} v_n(z) = 1 + u_1 z + u_2 z + \ldots + u_{n+1} z^{n+1} &= \sigma_n \sum_{k=0}^{n+1} \phi_k(0)^* \phi_k(z) \\ &= \frac{\phi_{(n+1)*}}{\phi_{n+1,0}} \end{aligned}$$

and σ_n is the minimum prediction error given by:

$$\sigma_n = \frac{1}{\sum_{k=0}^{n+1} |\phi_k(0)|^2}.$$

Therefore, by the Kolmogorov isomorphism $\hat{X}_{k+1|n}$, denoting the best linear least squares predictor of X_{k+1} given $X_k, X_{k-1}, \ldots, X_{k-n}$, which solves Problem 2.5.1 can be written as:

$$\begin{aligned}\hat{X}_{k+1|n} &= -\sum_{j=0}^{n} u_{j+1}^* X_{k-j} \\ &= -\int_{-\pi}^{\pi} e^{i\lambda}(v_n(e^{i\lambda})^* - 1) dZ_X(\lambda).\end{aligned}$$

Assume that F_X is absolutely continuous with spectral density f_X (hence X is a PND process). Then it can be shown that as $n \uparrow \infty$, $\hat{X}_{k+1|n}$ converges in $H(X)$ to the optimal one step ahead predictor $\hat{X}_{k+1}$ given by:

$$\hat{X}_{k+1} = \int_{-\pi}^{\pi} e^{i\lambda} \left(1 - \frac{g(0)}{g(e^{i\lambda})^*}\right) dZ_X(\lambda),$$

where g is as defined before. The formula above gives an explicit relationship between spectral factorization and best linear one step ahead prediction filters.

In engineering, the function g gives rise to two important filters. First of all, $g_*(z)$ has the interpretation as a *modelling filter* for the process X. This means that $g_*(z)$ is a causal and stable transfer function of a linear time invariant filter which generates the stationary process X when it is driven by a white noise sequence. Secondly, if $g_*(z)$ does not have any roots on the unit circle then $\frac{1}{g_*(z)}$ has the interpretation as the causal and stable transfer function of an *innovation filter* for X. The innovation filter outputs a white noise sequence with unit variance when driven by the process X_k. For details on time invariant filters for WSS processes, see, e.g., [34, Section 4.10].

Summarizing our discussion so far, we have now seen the close relationship that exists between prediction theory and spectral factorization. This relationship between predictors and spectral factors have been exploited by researchers for constructing prediction and innovation filters from spectral factors (this is usually attributed to N. Wiener who originally proposed spectral factorization [44] as an ingenious method for solving a so-called Wiener-Hopf integral equation in linear estimation theory), and, conversely, to construct spectral factors from the innovation filters as in [35, 36, 37]. Kolmogorov's work on prediction theory of scalar WSS processes was extended to multivariate/vector WSS processes by Wiener and Masani [39, 40, 41], and also, independently of Wiener and Masani, by Helson and Lowdenslager [42]. Although the analysis becomes much more

complicated, most of the results we have discussed have an extension to the vector case. A similar remark is also true for Szegö's work on orthogonal polynomials, it has been extended to the vector case by researchers in statistics and circuit theory (see [45, 11, 10]).

2.6 ARMA models, Darlington synthesis and generalization of the Szegö-Levinson algorithm

A shortcoming of using $\frac{1}{\phi_{n*}}$ as an approximate modelling filter for $\{X_k\}_{k\in\mathbb{Z}}$ is that it is an *all-pole/autoregressive* (AR) filter. Thus, its frequency response tends to have a "flat" profile for small and medium values of n. It is desirable to have autoregressive moving average (ARMA) modelling filters which have rational transfer functions with possibly some zeros inside or on the unit circle. The theory of orthogonal polynomials does not provide obvious insights into how the prediction theory can be modified to obtain ARMA filters. It was a surprising connection between the Szegö-Levinson algorithm and a certain *Darlington synthesis* procedure in circuit theory, first reported in [8], which allowed the extension to be realized. The connection is that the Szegö-Levinson algorithm may be viewed as a special case of Darlington synthesis, i.e., it corresponds to Darlington synthesis with extraction of "sections" whose transmission zeros are all at the origin (see [8]). Since the Darlington synthesis is actually a generalized version of the classical Schur algorithm/recursion [46, pp. 101-104] applied to the graph symbol of a "passive scattering function", extension of the Szegö-Levinson algorithm is facilitated by simply extracting sections corresponding to transmission zeros which are not at the origin. Following [37], by passive scattering function we mean a scalar function which is analytic on $\mathbb{D}$ (or on $\mathbb{C}\backslash\overline{\mathbb{D}}$ as in [35], depending on the setting on which one is considering the problem. This causes no confusion, as long as one works consistently on either $\mathbb{D}$ or $\mathbb{C}\backslash\overline{\mathbb{D}}$) and is bounded there in magnitude by 1. They are also known as *Schur functions* and we shall denote the class of such functions by $\mathcal{S}$. The Schur recursion (from this point onwards when referring to the Schur recursion or algorithm we implicitly mean the generalized version discussed in [8, 35, 37]) allows sections to be extracted in a sequential manner, one after another, with one section corresponding to exactly one transmission zero. In the development of the theory, an important role is played by

the so-called J-lossless matrices. Indeed, each section will in fact be a J-lossless matrix. In the scalar theory developed in [35, 37], these are complex $\mathbb{C}^{2\times 2}$-valued functions Θ which satisfy:

1. Each element of Θ belongs to the Nevanlinna class of complex functions on $\mathbb{D}$ which can be written as the ratio $\frac{f}{g}$ of two complex functions f, g which are analytic and bounded on $\mathbb{D}$.

2. Θ is J-contractive on $\mathbb{D}$: $\Theta(z)J\overline{\Theta(z)} \leq J$ for almost all $z \in \mathbb{D}$.

3. Θ is J-unitary on $\mathbb{T}$: $\Theta(z)J\overline{\Theta(z)} = J$ for almost all $z \in \mathbb{T}$,

where J is the matrix:

$$J = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix},$$

and $\overline{\Theta}$ denotes the elementwise complex conjugation of Θ.

Some important properties of J-lossless matrices are listed in [35, Theorem 2.1]. Although the key ideas for the extension are given in [8], their full exploitation and comprehensive treatment of the ideas, including the connection with (generalized) prediction of WSS processes and the related convergence results, were given in [35] for the scalar case and in [36] for the vector case. The papers [35, 36] deal with extraction of sections with a transmission zero inside $\mathbb{D}$ but not on $\mathbb{T}$. Unlike the Szegö-Levinson algorithm, which outputs a polynomial innovation filter (with no poles), the Schur algorithm produces a rational innovation filter with pre-specified poles which coincide with the zeros of the extracted sections. Therefore, the associated modelling filter will be rational having zeros which coincide with transmission zeros of the sections. An exposition of how sections with a transmission zero on $\mathbb{T}$ can be extracted was first given in [47], but without any proofs. The theoretical analysis appears later in the paper [37]. The context of [37] is actually to solve what is known as the *lossless inverse scattering* (LIS) problem, but as a bonus, thanks to a special embedding property of J-lossless matrices, one obtains a solution to Problem 2.6.1, which is a generalized version of Problem 2.5.2, and an approximate rational spectral factor of the associated spectral density. However, there are limitations of this approach for the purpose of spectral factorization and this will be discussed in the next section.

Problem 2.6.1 *Let F_X be the spectral distribution of a non-deterministic WSS process. Given a positive integer n and a pre-specified polynomial h_n of degree at*

most n such that all its roots are in $\mathbb{C}\backslash\mathbb{D}$ and $\int_{-\pi}^{\pi} \frac{1}{|h_n(e^{i\lambda})|^2} dF(\lambda) < \infty$, find the unique polynomial $a(z) = a_{n,0}z^n + a_{n,1}z^{n-1} + \ldots + a_{n,n}$ of degree at most n which minimizes the integral

$$\int_{-\pi}^{\pi} \left| 1 - \frac{a(e^{i\lambda})}{h_n(e^{i\lambda})} \right|^2 dF_X(\lambda)$$

For the remainder of the discussion, we assume that the associated WSS process is PND with a spectral density W defined on $\mathbb{T}$ (previously we have defined spectral densities on $(-\pi, \pi]$, but this causes no difficulty since the map $\lambda \in (-\pi, \pi] \mapsto e^{i\lambda} \in \mathbb{T}$ is bijective). Then one may associate a unique function $Z \in \mathcal{S}$, referred to as the impedance function, to W such that $Z_* + Z = W$ on $\mathbb{T}$. The impedance function is related to W via the *Herglotz representation*:

$$Z(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\lambda} + z}{e^{i\lambda} - z} W(e^{i\lambda}) d\lambda \quad \forall z \in \mathbb{D}, \tag{2.5}$$

and

$$Z(e^{i\lambda}) = \lim_{r \uparrow 1} Z(re^{i\lambda}) \text{ a.e. } \mathbb{T}.$$

In general, the right hand side of (2.5) cannot always be integrated explicitly. Therefore, typically Z would have be evaluated numerically. The reason we have introduced Z is because the Schur algorithm is actually applied to the passive scattering function S defined by $S = \frac{Z-1}{Z+1}$. Thus, to apply the algorithm for a given spectral density W, S must first be constructed via Z. We shall now also introduce, following [37], the notion of *point of local losslessness* (PLL). Formally, a point $b \in \mathbb{T}$ is a PLL of order k if

$$\int_{-\pi}^{\pi} \frac{W(e^{i\lambda})}{|e^{i\lambda} - b|^{2k}} d\lambda < \infty$$

and

$$\int_{-\pi}^{\pi} \frac{W(e^{i\lambda})}{|e^{i\lambda} - b|^{2k+2}} d\lambda = \infty,$$

for an integer $k \geq 1$. In the Schur algorithm described in [37] there are two types of sections which can be extracted, depending on the location of the associated transmission zero:

1. If the transmission zero is chosen in $\mathbb{D}$ then we may extract what is called a *Schur section* which is analytic on $\overline{\mathbb{D}}$.

2. If the transmission zero is a PLL then one may extract a *Brune section* which has a pole at a point on $\mathbb{T}$.

The Schur algorithm/Darlington synthesis procedure goes as follows. Given $S \in \mathcal{S}$ as defined above which has been normalized such that $S(0) = 0$ (note that this is equivalent to normalizing W such that $\frac{1}{2\pi}\int_{-\pi}^{\pi} W(e^{i\lambda})d\lambda = 1$), we form the vector $[\ 1 \quad S_0\ ]^\top$ with $S_0 = S$. Then depending on whether the transmission zero z_1 is in $\mathbb{D}$ or on $\mathbb{T}$, we may extract a Schur or Brune section θ_i (which is a J-lossless matrix) to get $[\ A_1 \quad B_1\ ]^\top = \theta_i[\ 1 \quad S_0\ ]^\top$. Defining $S_1 = \frac{B_1}{A_1}$ then S_1 will again be in $\mathcal{S}$ and satisfies $S_1(0) = 0$. Replacing S_0 with S_1 the procedure may be iterated, by choosing $z_2, z_3, \ldots$ to obtain $\{\theta_2, A_2, B_2, S_2\}, \{\theta_3, A_3, B_3, S_3\}, \ldots$, and so on. Letting $\Theta_n = \theta_n\theta_{n-1}\cdots\theta_1$, then Θ_n will be J-lossless and has the structure:

$$\Theta_n = \begin{bmatrix} R_{n*}^{-1}(1+Z_n) & R_{n*}^{-1}(1-Z_n) \\ F_n^{-1}(1-Z_n) & F_n^{-1}(1+Z_n) \end{bmatrix}, \tag{2.6}$$

where $F_{n*}F_n = R_nR_{n*} = Z_n + Z_{n*}$. Note that one may recover F_n and R_n (which are both rational functions) from Θ_n using the relations $F_n = (\Theta_{n,21} + \Theta_{n,22})^{-1}$ and $R_n = (\Theta_{n,11*} + \Theta_{n,12*})^{-1}$. It can then be shown [35] that if one lets $n \to \infty$ and chooses $z_1, z_2, \ldots \in \mathbb{D}$ (hence only Schur sections are extracted) such that it satisfies $\sum_{n=0}^{\infty}(1 - |z_n|) = \infty$, F_{n*}^{-1} converges to an innovation filter for a WSS process with spectral density W, while F_n converges to a spectral factor of W (i.e., a modelling filter for the process) in the sense that:

$$\lim_{n\to\infty}\int_{-\pi}^{\pi} |F_{n*}(e^{i\lambda})^{-1} - F_*(e^{i\lambda})^{-1}|^2 W(e^{i\lambda})d\lambda = 0.$$

If some Brune sections are also extracted (corresponding to a transmission zero on $\mathbb{T}$), no convergence results were given in [37], instead some explicit formulas for the approximation error $\int_{-\pi}^{\pi} |F_{n*}(e^{i\lambda})^{-1} - F_*(e^{i\lambda})^{-1}|^2 W(e^{i\lambda})d\lambda$ are derived. In this case F_{n*} and $\frac{1}{F_{n*}}$ can still serve as useful approximate rational modelling and innovation filters, respectively, if the transmission zeros are chosen appropriately as to keep the approximation error small.

2.7 Limitations of the Schur algorithm

The Schur algorithm, although elegant and results from a beautiful combination of theoretical insights from circuit theory and harmonic analysis, has some limitations which we shall elaborate upon in this section.

It has long been observed that the Szegö-Levinson algorithm converges slowly when W is *non-coercive* (has zeros on the unit circle) or is almost non-coercive.

This is also the case with all other spectral factorization algorithms that the author is aware of. In the case of the Szegö-Levinson and assuming that W is rational, an asymptotic analysis of this behavior is given in [12, 9] and goes as follows. Let us write W as $W(e^{i\theta}) = \frac{a_*(e^{i\theta})a(e^{i\theta})}{b_*(e^{i\theta})b(e^{i\theta})}$, where a and b are co-prime polynomials having no roots in $\mathbb{C}\backslash\overline{\mathbb{D}}$ with $a(0) = 1$. Let $c_0, c_1, \ldots$ be the covariance sequence associated with W and let $r_0, r_1, \ldots$ be the (unique) corresponding *Schur parameters* (for details see [7, 34, 12, 9]). Then in [12, 9] it was shown that a certain *almost recurrence* relation on the sequence $r_k, r_{k+1}, \ldots$ can be established for all k large enough. In particular, $r_k, r_{k+1}, \ldots$ is *almost rational*, and a will be an *almost recurrence polynomial*, in the terminology of [12, 48], for the sequence. To be precise, if $a = 1+a_1z+\ldots a_lz^l$ ($a_l \neq 0$) then for any given $\epsilon > 0$ there exists an integer $K(\epsilon)$ such that $r_k, r_{k+1}, \ldots, r_{k+l}$ obeys the almost recurrence relation:

$$|r_{k+1} + a_1r_{k+2} + \ldots + a_lr_{k+l}| < \epsilon \max_{k<m<k+l}\{|r_m|\},$$

for all $k > K(\epsilon)$ (note that Schur parameters satisfy $|r_k| \leq 1$ for all $k \geq 0$).

A consequence of the almost recurrence is that the asymptotic rate of decay of the Schur parameters are dictated by the location of zeros of a. If the roots of a are away from $\mathbb{T}$, the rate of decay is geometric, but the rate steadily decreases, although still geometric, as the location of the zeros come closer to the unit circle. When there is a root on $\mathbb{T}$ then the geometric rate is lost. Since it is well known that $\sigma_n = \Pi_{k=0}^n(1 - |r_k|^2)$ (σ_n is as defined in Section 2.5, see for example [12]), we clearly see that the Szegö-Levinson algorithm converges at the same rate as the rate of decay of r_k. Hence, when a has roots close to or on the boundary, r_k decays slowly and the Szegö-Levinson algorithm follows suit.

The same line of reasoning as given above also applies to the Schur algorithm by using an interpretation of that algorithm given by Delsarte and Genin in [38]. The latter interpretation allows us to see more lucidly the connections of the Schur algorithm to the Szegö-Levinson in terms of Toeplitz matrices. Suppose that the transmission zeros $z_1, z_2, \ldots \in \overline{\mathbb{D}}$ are chosen such that $z_k = 0$ for all k large enough (thus $\sum_{k=1}^{\infty}(1 - |z_k|) = \infty$). Let us define $h_0 = 1$ and $h_n(z) = \Pi_{k=1}^n(1 - z_k^*z)$ for $n \geq 1$, and let $W_n = \frac{W}{|h_n|^2}$. Suppose also that W is a bounded rational spectral density and W_n is integrable on $\mathbb{T}$ for all n (this will be the case if any z_k on $\mathbb{T}$ is a PLL of W and the multiplicity of z_k does not exceed the order of the PLL). Denote the covariance sequence associated with W_n by $c_{n0}, c_{n1}, \ldots$, i.e., $c_{nk} = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_n(e^{i\lambda})e^{-ik\lambda}d\lambda$. Then the family of finite sequences $c_{n0}, c_{n1}, \ldots, c_{nn}$ for $n = 0, 1, 2, \ldots$ form what is referred to in [38] as a *first degree*

Toeplitz family. To see this, we note that $W_k = \frac{W_{k-1}}{|z-z_{k-1}|^2}$ for all $k \geq 1$ with $W_0 = W$. Then from [38, Eq. (7)] it is easy to see that the collection of partial sequences $c_{k0}, c_{k1}, \ldots, c_{kk}$, $k = 0, 1, \ldots, n$, will generate such a family of Toeplitz matrices.

Let $\phi_{n1}, \ldots, \phi_{nn}$ be the first n sequence of orthogonal polynomials with respect to the spectral density W_n (thus ϕ_{nn} may be computed with the Szegö-Levinson algorithm). Thus, for fixed n, $\frac{\phi_{nn*}}{\phi_{nn,0}}$ (here $\phi_{nn,0}$ denotes the leading coefficient of ϕ_{nn}) is the solution to Problem 2.5.2 given a WSS process $\{X_{nk}\}_{k\in\mathbb{Z}}$ with spectral density W_n or, equivalently, with covariance sequence $c_{n0}, c_{n1}, \ldots$, for all n. As shown in [38], $\phi_{00}, \phi_{11}, \phi_{22}, \ldots$ can be computed recursively, and this is exactly what the Schur algorithm in [37] actually does. Therefore, $\frac{1}{\phi_{nn*}}$ is an approximate modelling filter for $\{X_{nk}\}_{k\in\mathbb{Z}}$ while $\frac{h_{n*}}{\phi_{n*}}$ is an approximate modelling filter for a process $\{X_k\}_{k\in\mathbb{Z}}$ with spectral density W. In fact, $\frac{h_n}{\phi_{nn}} = F_n$ [37, pp. 654-656], where F_n is as defined in Section 2.6. However, if W has roots almost at or on $\mathbb{T}$ which are not *exactly cancelled* by a corresponding root of h_n then, by the almost recurrence property of the Schur parameters that we have just discussed, F_n converges slowly to a spectral factor of W. The same reasoning applies to the general case where z_k is not necessarily zero for all k large enough, since for any *fixed* n which is not large, the presence of uncancelled zeros of W in W_n implies that ϕ_{nk} can only be a "good" approximate innovation filter for k much larger than n.

In practice, exact cancellation of roots of W on $\mathbb{T}$ is of course not possible to achieve due to approximation and numerical errors. However, this is not the only problem that can be encountered. A more fundamental limitation is that in some instances W may have a zero $b \in \mathbb{T}$ which is *not* PLL, i.e., there does not exist any positive integer k such that $\int_{-\pi}^{\pi} \frac{W(e^{i\lambda})}{|e^{i\lambda}-b|^{2k}} d\lambda < \infty$. An example of such a spectral density is the following:

$$W(e^{i\lambda}) = \sqrt{\frac{1+\cos\lambda}{1+\cos\lambda+\sigma(1-\cos\lambda)}}, \quad \sigma > 0 \tag{2.7}$$

which has a zero at $b = -1$ but which may be inspected not to be a PLL. Hence it is not possible to extract a Brune section with transmission zero at $b = -1$ in order to "cancel" that particular zero of W and accelerate convergence.

Therefore, we see that there are circumstances where the Schur algorithm is inadequate for numerical spectral factorization. However, to the best of our knowledge the Schur algorithm (of which the Szegö-Levinson algorithm is a special case) seems to be the only algorithm which allows placement of desired zeros on

$\overline{\mathbb{D}}$ and applies to a reasonably large class of spectral densities, including non-rational and non-coercive ones. For this reason, in Chapter 6 we propose a new approach to spectral factorization which can be applied to derive approximate rational spectral factors of spectral densities such as of the type given in (2.7) with zeros on the unit circle which need not be PLL.

Chapter 3

Results on Bounded Solutions of the Rational Covariance Extension Problem

3.1 Introduction

In the last chapter we have explained the motivation of our work in the context of wide sense stationary processes, and reviewed some related concepts, including the notion of spectral factorization. In this chapter we shall describe the so-called rational covariance extension problem with degree constraint (RCEP). Our interest in the RCEP is as a tool for developing a new approach to spectral factorization that will be proposed in Chapter 6 of the thesis.

Recent years have seen significant advances in the theory of analytic interpolation on the open unit disc of the complex plane. Some major results are the parametrization of all positive real rational functions interpolating a certain positive partial covariance sequence $c_0, c_1, \ldots, c_n$, in terms of desired "spectral zeros" and the introduction of a convex optimization based approach to compute the solution [9, 16, 5, 48, 6]. However, the convex optimization approach was originally developed for the case where none of the spectral zeros lie on the unit circle. The remaining case where there are spectral zeros on the unit circle is important not only for the sake of completeness, but also due to the fact that placing or forcing a zero on the unit circle is desirable, such as in the design of some filters. In this chapter, we derive some new theoretical results for this special case based on convex optimization. An alternative treatment based on solving non-linear equations has been given in [49]. However, there are important new insights gained with the

current approach. For example, we are able to derive a necessary and sufficient condition for a solution to be bounded (have no poles on the unit circle). We also assert, and demonstrate by numerical examples, that bounded solutions can be computed using methods that have been developed for pseudopolynomials free of zeros on the unit circle. In fact, building on the ideas developed in this chapter and the next, we shall show in Chapter 5 that an earlier homotopy continuation algorithm due to Enqvist [50] can compute all rational covariance extensions of a bounded degree. This was *not* previously known and could be advantageous in view of the current lack of theoretical convergence results for the algorithm of [49] and the more general nature of the algorithm of [50] (to be discussed further in Chapter 5).

More recently in [51], a theory of generalized interpolation with complexity constraint has emerged as an extensive generalization of the convex optimization approach first presented in [5]. The focus of [51] is on theoretical development (rather than numerical development as in [49]) and applies to a general, possibly abstract, class of interpolation problems with complexity constraint (a generalization of the notion of degree constraint). In particular, it also covers the case where the parametrizing pseudopolynomial has zeros on the unit circle. Our analysis, which is also based on convex optimization, proceeds in a different manner from [51]. Intrinsic and important differences between our work and [51] will be discussed. In particular, we argue that our results do not follow obviously from [51].

The discussion of this chapter is adapted from the papers [17, 52] (joint work with A. Bagchi).

3.2 The rational covariance extension problem (RCEP)

In this section we shall formally define the rational covariance extension problem (RCEP).

Definition 3.2.1 *A sequence of complex numbers $c_0, c_1, \ldots, c_n$ (with $c_0 \in \mathbb{R}$) is said to be a partial covariance sequence (PCS) if the Toeplitz matrix $T = [c_{j-i}]_{i,j=1}^{n+1}$, with $c_{-|i|} = c_{|i|}^*$, is positive definite.*

Problem 3.2.2 (RCEP) *Given a PCS $c_0, c_1, \ldots, c_n$ ($n \geq 1$), find all rational functions $f \in \mathcal{C}$ of McMillan degree less than or equal to n such that the first*

$n+1$ coefficients of the Taylor series expansion of f about 0 is $\frac{1}{2}c_0, c_1, \ldots, c_n$.

The RCEP basically adds a new requirement of degree bound to the classical Carathéodory extension problem which is traditionally solved by Schur's algorithm [12]. A drawback of Schur's algorithm is that, in general, it does not give a convenient parametrization of solutions of a bounded degree. The Carathédory extension problem is related to the classical Nevanlinna-Pick interpolation problem which was solved by Nevanlinna by an algorithm similar to Schur's [53], sometimes known as the Nevanlinna-Schur algorithm.

In a series of papers [9, 16, 48], a complete parametrization of all solutions of the RCEP has been established. We now state a pertinent result:

Theorem 3.2.3 *For a given PCS and any polynomial $\eta \neq 0$ of degree $\leq n$ with roots in $\mathbb{C}\backslash\mathbb{D}$ and normalized by $\eta(0) = 1$, there exists a unique pair of polynomials (π, χ) of degree $\leq n$ such that $\chi(0) > 0$, $\pi + \chi$ has all its roots in $\mathbb{C}\backslash\mathbb{D}$, the pair satisfies the relation*

$$\pi\chi_* + \chi\pi_* = \kappa^2\eta\eta_* \tag{3.1}$$

for a fixed $\kappa > 0$, and $f = \frac{\pi}{\chi}$ satisfies the requirements of the RCEP.

Remark 3.2.4 *This theorem is stated slightly differently from [48, Theorem 2]. We have added the requirement $\chi(0) > 0$ and κ fixed so that the pair (π, χ) is unique. In [48], it is implicit that the uniqueness of (π, χ) is in the equivalence class of graph symbols.*

The parametrization given in the theorem may also be stated equivalently in terms of so-called *pseudopolynomials*. By a pseudopolynomial we mean a complex function of the form $f(z) = a_0 + \sum_{k=1}^{n} \left(a_k^* z^{-k} + a_k z^k\right)$, where $0 \leq n < \infty$, $a_n \neq 0$ and $(a_0, a_1, \ldots, a_n) \in \mathbb{R} \times \mathbb{C}^n$. Then n is said to be the *order* or *degree* of the pseudopolynomial f (the order is zero if f is a constant function). Let $\mathfrak{Q}(n, A)$ denote the set of all pseudopolynomials of order *at most* n with $(a_0, a_1, \ldots, a_n) \in \mathbb{R}\times A^n$ where $A \subseteq \mathbb{C}$. We induce a topology on this set by the maximum norm: $\|f\|_\infty = \max_{z\in\mathbb{T}} |f(z)|$. We also define $\mathfrak{Q}_+(n, A)$ to be the set of all elements of $\mathfrak{Q}(n, A)$ which are strictly positive (> 0) on $\mathbb{T}$. With pseudpolynomials having been defined, we may equivalently state the parametrization of all solutions of the RCEP in Theorem 3.2.3 in terms of elements $d \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$, where $d = \kappa^2\eta\eta_*$, and we can state a more specific problem, the particular rational covariance extension problem (PRCEP):

Problem 3.2.5 (PRCEP) *Given a PCS* $c_0, c_1, \ldots, c_n$ $(n \geq 1)$ *and a pseudopolynomial* $\Psi \in \overline{\mathfrak{Q}_+(n,\mathbb{C})} \backslash \{0\}$*, find the rational function* $f = \frac{a}{b} \in \mathcal{C}$ *of McMillan degree* $\leq n$ *such that the first* $n+1$ *coefficients of the Taylor series expansion of* f *about* 0 *is* $\frac{1}{2}c_0, c_1, \ldots, c_n$ *and* $ab_* + ba_* = \Psi$.

A convex optimization method for computing solutions of the PRCEP for any given real valued PCS $c_0, c_1, \ldots, c_n$ and pseudopolynomial $\Psi \in \mathfrak{Q}_+(n,\mathbb{R})$ (i.e., Ψ is free of roots on $\mathbb{T}$) was first given in [5, 6], and was subsequently adapted to solve the Nevanlinna-Pick interpolation problem with degree constraint in [14]. However, a specialized aspect of the theory which has received relatively less attention is the case of solving the PRCEP when the pseudopolynomial has zeros on the boundary. In this work, we extend the method of [5, 6] to the case where the pseudopolynomial has zeros on the boundary. It turns out that this leads to interesting new theoretical insights, including a necessary and sufficient condition for a $\mathcal{H}^\infty$ solution, as shown in the next section. A numerical treatment of the problem was recently given in [49] based on solving non-linear equations. There the orientation is towards computation of any real solution of the RCEP.

3.3 Main results

In this section we derive some properties of the solutions of the RCEP when the parametrizing pseudopolynomial has zeros on $\mathbb{T}$. In particular we show a necessary and sufficient condition for a solution to be in $\mathcal{H}^\infty$ and establish sequential continuity of the map from Ψ to the minimizer of a certain functional $\mathbb{J}_\Psi$ (to be defined below).

Define the mapping $Q : \mathbb{R} \times \mathbb{C}^n \to \mathfrak{Q}(n,\mathbb{C})$ by:

$$Q(q_0, q_1, q_2, ..., q_n)(z) = q_0 + \sum_{k=1}^{n} \frac{1}{2}(q_k^* z^{-k} + q_k z^k). \tag{3.2}$$

Clearly Q is a bijection.

Remark 3.3.1 *For shorthand, we shall write the integral* $\frac{1}{2\pi}\int_{-\pi}^{\pi} f\left(e^{i\theta}\right) g\left(e^{i\theta}\right)^* d\theta$ *as* $\langle f, g\rangle$.

For any $\Psi \in \overline{\mathfrak{Q}_+(n,\mathbb{C})} \backslash \{0\}$ we consider the functional $\mathbb{J}_\Psi : \overline{Q^{-1}\left(\mathfrak{Q}_+(n,\mathbb{C})\right)} \to \mathbb{R} \cup \{\infty\}$ defined by:

$$\mathbb{J}_\Psi(q) = \Re\big\{c^* q - \langle \Psi, \log Q(q)\rangle\big\}, \tag{3.3}$$

where $c = \mathrm{col}(c_0, c_1, \ldots, c_n)$ and $q = \mathrm{col}(q_0, q_1, \ldots, q_n)$. Here $\mathrm{col}(a_1, a_2, \ldots, a_n)$ denotes the column vector $[a_0\, a_1\, \ldots\, a_n]^T$.

Note that $\mathbb{J}_\Psi$ can be viewed as an extension to $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ of the functional φ that was defined in [6, (4.1)] for the special case where $\Psi \in \mathfrak{Q}_+(n,\mathbb{R})$ and $c_0, c_1, \ldots, c_n$ is real-valued. It then follows by close inspection of the proofs that certain key results in [6] can be easily extended to the current setting where $c_0, c_1, \ldots, c_n$ is complex-valued and $\Psi \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$. In particular, we state the analogues of Lemma 4.2, Lemma 4.3 and Proposition 4.6 of [6] in the following theorem:

Theorem 3.3.2 *$\mathbb{J}_\Psi$ has the following properties for any $\Psi \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$:*

- *$\mathbb{J}_\Psi$ is finite and continuous at any $q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$, except at zero. The functional is infinite, but continuous, at $q = 0$. Moreover, $\mathbb{J}_\Psi((1-t)q_0+tq_1)$ is a C^∞ function w.r.t. t for any $q_0, q_1 \in Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$.*
- *$\mathbb{J}_\Psi$ is strictly convex on the closed, convex domain $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$.*
- *For all $r \in \mathbb{R}$, $\mathbb{J}_\Psi^{-1}(-\infty, r]$ is compact. Thus $\mathbb{J}_\Psi$ is proper (i.e., $\mathbb{J}_\Psi^{-1}(A)$ is compact whenever A is compact) and bounded from below.*
- *The functional $\mathbb{J}_\Psi$ has a unique minimum on $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$.*

We now state the first result on a solution of the RCEP corresponding to a pseudopolynomial having zeros on $\mathbb{T}$:

Theorem 3.3.3 *If $q_{\min} \in Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$ is a minimum for $\mathbb{J}_\Psi$ then the solution of the PRCEP is: $f = \frac{a}{b}$ where $bb_* = Q(q_{\min})$ and $ab_* + ba_* = \Psi$. Conversely, suppose that $f = \frac{a}{b}$ is the solution to the PRCEP with b being an antistable polynomial (i.e., having roots strictly in $\mathbb{C}\backslash\overline{\mathbb{D}}$) and $ab_* + ba_* = \Psi$. Then $q_{\min} = Q^{-1}(bb_*)$ is a unique minimum for $\mathbb{J}_\Psi$.*

Proof. By inspection of the proofs of [6, Theorems 4.7 and 4.8] and using the directional derivative to replace the ordinary derivative, it follows those proofs remain valid if the polynomial $\sigma = z^n + \sigma_1 z^{n-1} + \ldots + \sigma_{n-1}z + \sigma_n$ of degree n defined in equation (2.18) of [6] is complex and not *Schur* (i.e., having roots in $\mathbb{D}$), but merely *stable* (i.e., having roots in $\overline{\mathbb{D}}$). Also note that $\sigma_*\sigma$ can be a pseudopolynomial of degree less than n if $\exists m$ satisfying $1 \le m \le n$, such that $\sigma_k = 0$ for all $k \ge m$. The main idea is that the minimizer of $\mathbb{J}_\Psi$ may be an interior point even when $\Psi = \sigma_*\sigma \in \partial\mathfrak{Q}_+(n,\mathbb{C}) \setminus \{0\}$. □

The minimizer of $\mathbb{J}_\Psi$ may then be found by a Newton descent type algorithm which has been outlined in [5, 6, 14]. We illustrate this in the following example.

Example 3.3.4 *Let the given partial covariance sequence be*

$$\{0.2115, 0.0728, -0.0396\}.$$

We choose the pseudopolynomial $\Psi(z) = z+2+z^{-1}$ *which has two zeros on the unit circle, i.e., both at* $z = -1$*, and seek a solution of the RCEP of degree* 2*. By using a Newton gradient descent algorithm we obtain* $q_{\min} = \operatorname{col}(8.6250, 3.5000, 2.0000)$*. It can be checked that* $q_{\min}$ *is in the interior of* $\overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{R}))}$*, and the solution of the PRCEP is*

$$f(z) = \frac{0.09877 + 0.1111z + 0.01234z^2}{8 + 2z - z^2}.$$

An interesting question now is: what could happen if the minimum of $\mathbb{J}_\Psi$ lies on the boundary of $Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))$? We first look at an insightful example.

Example 3.3.5 *Consider the Carathèodory function*

$$f(z) = \frac{1}{2}\frac{1 + \frac{z}{2}}{1 - \frac{z}{2}}. \tag{3.4}$$

The associated PCS is $1, \frac{1}{2}, \frac{1}{4}, \ldots$*. We choose the pseudopolynomial* $\Psi(z) = z + 2 + z^{-1}$ *having a double root at* $z = -1$*. By Newton gradient descent we find* $q_{\min} \approx \operatorname{col}(2, 0.66749, -1.3324)$*. The roots of* $Q(q_{\min})$ *are*

$$\{2.0013, -1.0061, -0.99396, 0.49967\},$$

and the approximate solution is

$$\hat{f} = \frac{0.3326 + 0.4978z + 0.1652z^2}{2.0135 + 0.9952z - z^2}.$$

Note how two roots of $Q(q_{\min})$ *are close to* $z = -1$*. Assuming that were it not for numerical discrepancies that both roots would be exactly* -1 *and cancel the two corresponding roots of* Ψ*, we find:* $\frac{\Psi(z)}{Q(q_{\min})(z)} = \frac{1.5001}{2.5010 - \left(z + \frac{1}{z}\right)}$ *which is the power spectral density of the Carathèodory function*

$$\hat{f}(z) = 0.49948\frac{1 + 0.4997z}{1 - 0.49967z},$$

a function close to the true function f *given in* (3.4)*. Observe that we have deliberately chosen* Ψ *such that* $q_{\min}$ *is intuitively expected to lie on the boundary,*

in contrary to Example 3.3.4 in which $q_{\min}$ is in the interior. To see this, note that f maybe written as $f = \frac{a}{b}$ with $a = (1+\frac{z}{2})(z+1)$ and $b = 2(1-\frac{z}{2})(z+1)$ so that $a_*b + b_*a = \Psi$ and b_*b share a common double root at $z = -1$. In fact, the purpose of this example is to illustrate a case where $q_{\min}$ is at the boundary and also seems to be a stationary point, and to motivate the next theorem. We shall consider this example again in Section V.

Remark 3.3.6 *When $q_{\min}$ is close to or on the boundary, numerical problems can arise when Newton descent is used to find $q_{\min}$. To improve the situation for $q_{\min}$ close to the boundary, the optimization problem can be reformulated and numerically solved by a continuation method [50]. In certain circumstances, the same also applies when $q_{\min}$ is at the boundary. This is discussed in Section 3.4.*

As it turns out, the generality of the observation in Example 3.3.5 can be formally proven. It is the content of the next theorem:

Theorem 3.3.7 *The solution of the PRCEP is in $\mathcal{H}^\infty$ if and only if $\mathbb{J}_\Psi$ has a stationary point in the interior or boundary of its domain. If $Q(q_{\min}) \in \partial\mathfrak{Q}_+(n,\mathbb{C})$ and $q_{\min}$ is stationary, then every root of $Q(q_{\min})$ on $\mathbb{T}$ will also be a root of Ψ on $\mathbb{T}$, and the solution of the PRCEP is of order less than n. In this case the solution is given by: $f = \frac{a}{b}$ where $bb_* = Q_+(q_{\min})$, $ab_* + b_*a = \tilde{\Psi}$, and*

1. *$Q_+(q_{\min}) \in \mathfrak{Q}_+(n,\mathbb{C})$ denotes the pseudopolynomial that is left behind after all factors $\left(z^{\pm 1} - e^{i\phi}\right)$ corresponding to the roots of $Q(q_{\min})$ on $\mathbb{T}$ have been removed from $Q(q_{\min})$.*

2. *$\tilde{\Psi}$ denotes the pseudopolynomial that is left behind after all factors $\left(z^{\pm 1} - e^{i\phi}\right)$ corresponding to the roots of $Q(q_{\min})$ on $\mathbb{T}$ have been removed from Ψ.*

Proof. We need only prove the initial statement that the solution of the PRCEP is bounded if and only if $\mathbb{J}_\Psi$ has a stationary point in the interior or boundary of its domain. The remaining statements of the theorem all follow from the proof of the initial statement. Let q be such that $Q(q) \in \partial\mathfrak{Q}_+(n,\mathbb{C})$ and such that all the roots of $Q(q)$ on $\mathbb{T}$ are also the roots of Ψ on $\mathbb{T}$. Let the set of all $q \in \mathbb{C}^{n+1}$ satisfying the previous two conditions be denoted by $\mathcal{M}_{n,\Psi}$. First we show that for any $q \in \mathcal{M}_{n,\Psi} \cup Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$, the directional derivatives of $\mathbb{J}_\Psi$ exist in all feasible directions. To this end, for any $q_0 \in Q^{-1}(\overline{\mathfrak{Q}_+(n,\mathbb{C})})$ we define the directional derivative:

$$\nabla\mathbb{J}_{q_0-q}(q) = \lim_{h\downarrow 0} \frac{\mathbb{J}_\Psi(q + h(q_0 - q)) - \mathbb{J}_\Psi(q)}{h}.$$

It is easy to check that if $q + h(q_0 - q) \in \partial\mathfrak{Q}_+(n,\mathbb{C})$ for all $0 \leq h < \zeta$ and some $\zeta > 0$, then $Q(q)$ and $Q(q_0)$ must share a root on $\mathbb{T}$. Since all roots of $Q(q)$ on $\mathbb{T}$ are also roots of Ψ on $\mathbb{T}$, it follows that $\frac{\Psi}{Q(q+h(q_0-q))}$ is uniformly bounded a.e. on $\mathbb{T}$ for all $q_0 \in Q^{-1}(\mathfrak{Q}(n,\mathbb{C}))$ and for all $h > 0$. From the mean-value theorem of calculus it follows that:

$$\begin{aligned}&\Psi(e^{i\theta})\frac{\log Q(q+h(q_0-q))(e^{i\theta}) - \log Q(q)(e^{i\theta})}{h}\\ &= \frac{\Psi(e^{i\theta})Q(q_0-q)(e^{i\theta})}{Q(q)(e^{i\theta}) + \eta(h,e^{i\theta})Q(q_0-q)(e^{i\theta})},\end{aligned}$$

where $0 < \eta(h,e^{i\theta}) < h$, for all θ except for a finite number for which $Q(q)(e^{i\theta}) = 0$. Since the right hand side of the last equality is uniformly bounded for almost all $(h,e^{i\theta}) \in [0,\frac{1}{2}] \times \mathbb{T}$, we have that $\lim_{h\downarrow 0}\langle \Psi, \frac{\log Q(q+h(q_0-q)) - \log Q(q)}{h}\rangle = \langle \Psi, \lim_{h\downarrow 0}\frac{\log Q(q+h(q_0-q)) - \log Q(q)}{h}\rangle$ by the Lebesque Dominated Convergence Theorem [54]. Therefore, for any $q = \mathrm{col}\,(q_0,\ldots,q_n) \in \mathcal{M}_{n,\Psi} \cup Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$ and any $q_0 = \mathrm{col}\,(q_{00},\ldots,q_{0n}) \in Q^{-1}(\overline{\mathfrak{Q}_+(n,\mathbb{C})})$ we get:

$$\begin{aligned}\nabla_{q_0-q}\mathbb{J}_\Psi\,(q) &= \Re\Big\{c^*\,(q_0-q) - \sum_{k=0}^{n}\langle\frac{\Psi}{Q(q)}, g_{k*}\rangle(q_{0k}-q_k)\Big\}\\ &= \Re\Big\{\sum_{k=0}^{n}\Big(c_k - \langle\frac{\Psi}{Q(q)}, g_k\rangle\Big)^*(q_{0k}-q_k)\Big\}.\end{aligned}$$

where $g_k(z) = z^k$.

Now we are ready to prove **necessity**. By Theorem 3.2.3 and since the solution of the PRCEP is bounded by hypothesis, we know that there is a unique $\Omega \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}$ such that $\langle\frac{\Psi}{\Omega}, g_k\rangle = c_k$, for $k = 0,1,\ldots,n$ and $Q^{-1}(\Omega)$ lies in $\mathcal{M}_{n,\Psi} \cup Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$. Setting $q = Q^{-1}(\Omega)$ then we have that $\nabla\mathbb{J}_{q_0-q}(q) = 0$. Hence that particular choice of q is a stationary point and it is the unique minimizer of $\mathbb{J}_\Psi$. This establishes the necessity.

We proceed to prove **sufficiency**. Let q be a stationary point of $\mathbb{J}_\Psi$ by letting $\nabla\mathbb{J}_{q_0-q}\,(q) = 0$ for all $q_0 \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$. Then $q \in \mathcal{M}_{n,\Psi} \cup Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$, otherwise $\nabla_{q_0-q}\mathbb{J}_\Psi(q) = +\infty\ \forall q_0 \in \mathfrak{Q}_+(n,\mathbb{C})$ (by the same arguments employed in the proof of [5, Lemma 5.4]), and we have:

$$\Re\Big\{\sum_{k=0}^{n}\Big(c_k - \langle\frac{\Psi}{Q(q)}, g_k\rangle\Big)^*(q_{0k}-q_k)\Big\} = 0 \tag{3.5}$$

Now, for any $q \in \mathcal{M}_{n,\Psi} \cup Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$ we may write $Q(q) = Q_+(q)Q_0\,(q)$ where $Q_0(q)$ is a pseudopolynomial with all its roots on $\mathbb{T}$ or is identically equal to 1

if no such roots exist, while $Q_+(q)$ is a pseudopolynomial which does not have roots on the boundary. Because all the roots of $Q(q)$ which are on the boundary are also roots of Ψ by hypothesis, we may write $\Psi = \tilde{\Psi}(q)Q_0(q)$, where $\tilde{\Psi}(q)$ is a pseudopolynomial defined by $\tilde{\Psi}(q) = \frac{\Psi}{Q_0(q)}$. After inserting the two identities into (3.5) we obtain:

$$\Re\Big\{\sum_{k=0}^{n}\left(c_k - \langle\frac{\tilde{\Psi}(q)}{Q_+(q)}, g_k\rangle\right)^{*}(q_{0k} - q_k)\Big\} = 0 \tag{3.6}$$

However, equation (3.6) holds for all $q_0 \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$. Therefore by inspection (e.g., see proof of [55, Lemma 5.1]) we must have

$$c_k - \langle\frac{\tilde{\Psi}(q)}{Q_+(q)}, g_k\rangle = 0 \Longleftrightarrow \langle\frac{\tilde{\Psi}(q)}{Q_+(q)}, g_k\rangle = c_k$$

for $k = 0, 1, \ldots, n$. Therefore, there is a unique Carathèodory function f such that $(f + f_*)(e^{i\theta}) = \frac{\tilde{\Psi}(q)(e^{i\theta})}{Q_+(q)(e^{i\theta})}$, f satisfies the interpolation constraints, and f is bounded. Hence we have shown sufficiency. Note the cancellation that takes place if $Q(q)$ has roots on the boundary. In this case the solution f will be of degree less than n. □

Therefore, stationarity of the minimizer of $\mathbb{J}_\Psi$ is essentially *a trademark for the boundedness of the solution*: if it is stationary then the solution is bounded, otherwise it is not. We may also show the following sequential continuity result:

Theorem 3.3.8 *Let* $\Psi \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$ *and let* $\{\Psi_k\}_{k\geq 1} \subset \overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$ *be a sequence such that* $\lim_{k\to\infty}\|\Psi - \Psi_k\|_\infty = 0$. *If*

$$q_{\min} = \underset{q\in\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}}{\arg\min}\mathbb{J}_\Psi(q) \quad \textit{and} \quad q_{\min,k} = \underset{q\in\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}}{\arg\min}\mathbb{J}_{\Psi_k}(q),$$

then

$$\lim_{k\to\infty}\|q_{\min} - q_{\min,k}\| = 0 \quad \textit{and} \quad \lim_{k\to\infty}\|Q(q_{\min}) - Q(q_{\min,k})\|_\infty = 0.$$

Proof. For $r > 0$, define the compact sets

$$B_r(q_{\min}) = \{q \in \mathbb{R}\times\mathbb{C}^n : \|q - q_{\min}\| \leq r\}$$

and $S_r(q_{\min}) = \partial B_r(q_{\min})$. Also define the compact sets $X_r(q_{\min}) = B_r(q_{\min}) \cap \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ and $Y_r(q_{\min}) = \partial X_r(q_{\min})$. We prove that given any $\epsilon > 0$ small enough such that $0 \notin X_\epsilon(q_{\min})$, there is a $K(\epsilon) \geq 1$ such that $q_{\min,k} \in B_\epsilon(q_{\min})\ \forall k > K(\epsilon)$. First, we observe that

$$\begin{aligned}|\mathbb{J}_\Psi(q) - \mathbb{J}_{\Psi_k}(q)| &\leq \langle|\Psi - \Psi_k|, |\log Q(q)|\rangle \\ &\leq \|\Psi - \Psi_k\|_\infty\langle\mathbf{1}, |\log Q(q)|\rangle \text{ if } q \neq 0,\end{aligned}$$

where $\mathbf{1}: z \mapsto 1\, \forall z \in \mathbb{T}$. If we define $D = \max_{q \in X_\epsilon(q_{\min})} \langle \mathbf{1}, |\log Q(q)| \rangle$, we have that $\forall q \in X_\epsilon(q_{\min})$:

$$|\mathbb{J}_\Psi(q) - \mathbb{J}_{\Psi_k}(q)| \leq D \left\| \Psi - \Psi_k \right\|_\infty$$

or more explicitly,

$$\mathbb{J}_\Psi(q) - D \left\| \Psi - \Psi_k \right\|_\infty \leq \mathbb{J}_{\Psi_k}(q) \leq \mathbb{J}_\Psi(q) + D \left\| \Psi - \Psi_k \right\|_\infty \tag{3.7}$$

For any $r > 0$, define $Z_r(q_{\min}) = Y_r(q_{\min})$ if $q_{\min} \in Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))$ and $Z_r(q_{\min}) = S_r(q_{\min}) \cap \overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))}$ if $q_{\min} \in \partial Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))$. Notice that $Z_r(q_{\min})$ is a compact set. Choose any $\epsilon > 0$ small enough such that $0 \notin X_\epsilon(q_{\min})$ and such that $Z_\epsilon(q_{\min}) \subset Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))$ if $q_{\min} \in Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))$. Next, for any $q \neq q_{\min}$ define the unit vector $u_q = \frac{q - q_{\min}}{\|q - q_{\min}\|}$, and for any $q \in Z_\epsilon(q_{\min})$ and any $0 < d < \epsilon$ define the functions $L_1(q, d) = \mathbb{J}_\Psi(q) - \mathbb{J}_\Psi(q_{\min} + d u_q)$ and $L_2(q, d) = \mathbb{J}_\Psi(q_{\min} + d u_q) - \mathbb{J}_\Psi(q_{\min})$. Clearly, from the strict convexity of $\mathbb{J}_\Psi$, $L_1(\cdot, d)$ and $L_2(\cdot, d)$ are continuous, positive-valued (> 0) functions on $Z_\epsilon(q_{\min})$. Furthermore, define $\delta_i(d) = \min_{q \in Z_\epsilon(q_{\min})} L_i(q, d)$ for $i = 1, 2$. Observe that $\delta_i(d) > 0$ for $i = 1, 2$, for if it is not then $\exists q \in Z_\epsilon(q_{\min})$ such that $L_1(q, d) = 0$ and/or $L_2(q, d) = 0$, contradicting the fact that they are positive-valued on $Z_\epsilon(q_{\min})$. Let us now choose a fixed $d \in (0, \epsilon)$. Choose $K_d(\epsilon)$ (note the dependence on d) large enough such that $\|\Psi - \Psi_k\|_\infty < \frac{\min\{\delta_1(d), \delta_2(d)\}}{3D}$ for all $k > K_d(\epsilon)$, then using (3.7) one easily gets that for any $q \in Z_\epsilon(q_{\min})$:

$$\mathbb{J}_{\Psi_k}(q_{\min}) < \mathbb{J}_{\Psi_k}(q_{\min} + d u_q) < \mathbb{J}_{\Psi_k}(q) \quad \forall k > K_d(\epsilon) \tag{3.8}$$

From (3.8) and the strict convexity of $\mathbb{J}_{\Psi_k}$ for all k, it follows that $q_{\min,k} \in X_\epsilon(q_{\min}) \backslash Z_\epsilon(q_{\min})$ for all $k > K_d(\epsilon)$.

Summarizing, we have shown that for every $\epsilon > 0$ such that $0 \notin X_{\frac{\epsilon}{2}}(q_{\min})$, $\exists K\left(\frac{\epsilon}{2}\right)$ such that for all $k > K\left(\frac{\epsilon}{2}\right)$, $q_{\min,k} \in X_{\frac{\epsilon}{2}}(q_{\min})$ is in the interior of $B_\epsilon(q_{\min})$, or in other words, $\lim_{k \to \infty} \|q_{\min} - q_{\min,k}\| = 0$. It follows immediately that $\lim_{k \to \infty} \|Q(q_{\min}) - Q(q_{\min,k})\|_\infty = 0$. This concludes the proof. □

Although one may view the last theorem as a corollary to [49, Theorem 3.1] when Ψ and the PCS $c_0, c_1, \ldots, c_n$ are real, it is an interesting result in its own right. Notice that its proof is based solely on properties of $\mathbb{J}_\Psi$ (see Theorem 3.3.2) and is independent of Theorem 3.2.3. On the other hand, [49, Theorem 3.1] was derived based on Theorem 3.2.3. In fact, we claim that it is possible to show the converse: Theorem 3.2.3 and [49, Theorem 3.1] can be derived using Theorems 3.3.2 and 3.3.8. This interesting ramification of Theorem 3.3.8 presents an alternative analysis of the RCEP, *including unbounded solutions*, which will be treated in Chapter 4.

3.4 Discussion, extensions, and application of results

Our convex optimization based approach is reminiscent of the extensive and abstract generalization of [5, 6] given in [51], but it may be inspected that the two treatments are not identical and there are two important differences which we shall now discuss.

First, the objectives of the two works are different. In [51], the objective is to extend the convex optimization technique to generalize Theorem 3.2.3 to the setting of a general class of interpolation problems with a so-called complexity constraint, whereas in the present work we do not attempt to re-derive Theorem 3.2.3, but rather to use the theorem and/or properties of $\mathbb{J}_\Psi$ when Ψ has zeros on $\mathbb{T}$ (to the best of our knowledge, we were the first to do this) to derive Theorems 3.3.3, 3.3.7 and 3.3.8. Secondly, our treatment is centered on analysis of boundary properties of the functional $\mathbb{J}_\Psi$ when Ψ may have zeros on $\mathbb{T}$. Although a generalized version of $\mathbb{J}_\Psi$ was formulated in [51], its properties when Ψ has zeros on $\mathbb{T}$ were not investigated. Instead, an alternative route was taken whereby the case $\Psi \in \partial\mathfrak{Q}_+(n,\mathbb{C})\backslash\{0\}$ is treated via analysis of a functional $\mathbb{K}_\Psi$ (see [51, eq. (2.16)]) defined on a set of Schur functions (i.e., functions which are analytic on $\mathbb{D}$ and bounded there in magnitude by one) satisfying a certain constraint. In particular, it has been shown that the unique extremal point of $\mathbb{K}_\Psi$ (which, in this case, is a maximizer) is always stationary (see the penultimate part of the proof of [51, Theorem 1] on uniqueness of a solution, p. 13). On the other hand, this is *not the case* for $\mathbb{J}_\Psi$. As we have shown, the extremal point of $\mathbb{J}_\Psi$ (which is a minimizer) *need not be stationary*. In fact, it is precisely this unique property of $\mathbb{J}_\Psi$ over $\mathbb{K}_\Psi$ which led us to a characterization of $\mathcal{H}^\infty$ solutions of the RCEP as stated in Theorem 3.3.7.

Continuing further, we note that for Ψ positive definite on $\mathbb{T}$, $\mathbb{K}_\Psi$ is obtained from a transformation of the functional $\mathbb{I}_\Psi$, the dual of $\mathbb{J}_\Psi$ (see [51, eq. (2.14)]). To derive our results within the development of [51], some results relating $\mathbb{K}_\Psi$ and $\mathbb{J}_\Psi$ need to be established for Ψ non-negative but not positive definite. Then one should show that the maximizer f of $\mathbb{K}_\Psi$ satisfies $\operatorname{ess\,inf}_{z\in\mathbb{T}} |1+f(z)| > 0$ (this is equivalent to the RCEP having a bounded solution) if and only if the minimizer of $\mathbb{J}_\Psi$ is stationary. These relations have not been considered in [51]. Thus, in light of these facts, our results do not obviously follow from [51]. On the contrary, it may be possible to generalize them to the setting of [51] by further analysis of

the generalized version of $\mathbb{J}_\Psi$. Indeed, we should keep in mind that our results are specialized to the RCEP, while those of [51] apply to a more general, possibly abstract, class of interpolation problems with a complexity constraint.

We now discuss some practical implications of Theorems 3.3.3 and 3.3.7. From Theorem 3.3.3 we see that when Ψ has zeros on $\mathbb{T}$ and the minimizer of $\mathbb{J}_\Psi$ is in the interior of $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ and away from the boundary, the solution can be computed rather quickly and easily by Newton descent. We have illustrated this in Example 3.3.4. When the minimizer is close to the boundary, the continuation method of [50] can be applied for good numerical results. For cases where Theorem 3.3.7 is applicable, it ought to also be possible to compute solutions by the continuation method. Example 3.3.5 indicates that even a standard Newton descent method can yield an approximate solution, albeit a crude one. Therefore, it is reasonable to expect the more robust continuation method to give good numerical results for such cases or for ones which are similar (i.e., almost cancellations of insignificant poles lying close to the boundary). Indeed, to support this claim we rework Example 3.3.5 using the continuation method:

Example 3.4.1 *Let c_0, c_1, c_2 and Ψ be as given in Example 3.3.5. Applying the continuation method with step length parameter ε=0.01 (see [50, p.1196]) yields $b(z)$=1.1547 + 0.5773z − 0.5774z^2, $a(z)$=0.5774 + 0.8660z + 0.2887z^2, and the corresponding solution is*

$$f(z)=\frac{0.5774+0.8660z+0.2887z^2}{1.1547+0.5773z-0.5774z^2}.$$

In fact, in Chapter 5 results will be developed that justify using the continuation method for computing not only the bounded solutions discussed in this chapter, but *all* solutions of the RCEP. Since convergence is better understood for that method, this can be beneficial because at present there are no theoretical convergence results for the alternative algorithm of [49]. Moreover, there are two other attractive features of the continuation method. First is that the Hessian of the modification of $\mathbb{J}_\Psi$ given in [50] can be inverted in a fast and efficient manner because of its special Toeplitz-plus-Hankel (T+H) structure [56, 57]. This kind of structure does not seem to be present in the latter algorithm. Secondly, it can be naturally extended to the setting of more general analytic interpolation and moment problems [58]. Further discussion of the continuation method will be postponed until Chapter 5. Thus, our results have extended the utility of the earlier methods of [6, 50].

3.5 Concluding remarks

The contributions of this chapter are some new theoretical results on solutions of the RCEP corresponding to $\Psi \in \partial\mathfrak{Q}_+(n,\mathbb{C})\backslash\{0\}$, i.e., the case where the parametrizing pseudopolynomial has zeros on $\mathbb{T}$. In particular, we show that for a solution to be in $\mathcal{H}^\infty$, it is necessary and sufficient that the minimizer of $\mathbb{J}_\Psi$ is stationary. Furthermore, we have shown that some solutions for this case can be computed using methods that have been developed for Ψ which is free of zeros on $\mathbb{T}$, extending the utility of those methods. We also establish the sequential continuity of a certain map based solely on the properties of $\mathbb{J}_\Psi$ and independently of the result on complete parametrization of all solutions of the RCEP (Theorem 3.2.3). Full exploitation of this result will be given in the next chapter.

We have also outlined the differences between our work and [51] which is also based on convex optimization but applies to a more general class of interpolation problems. We point out some interesting differences between the functionals $\mathbb{J}_\Psi$ and $\mathbb{K}_\Psi$, which are the main object of the analysis of, respectively, this chapter and [51], and argue that our results do not obviously follow from [51] and that it may be possible to generalize them to the setting of [51].

Note that although this chapter specifically treats the RCEP, the results presented here readily extends to the Nevanlinna-Pick interpolation with degree constraint as described in [14, 59, 60]. This more general setting will be taken up in Chapter 5 when we further develop the method of [50] as a tool for computing all degree constrained rational interpolants.

Chapter 4

Results on General Solutions of the Rational Covariance Extension Problem

4.1 Introduction

In this chapter, we continue to develop the ideas of Chapter 3, in which $\mathcal{H}^\infty$ solutions of the RCEP were studied, and derive new results relating to *all* solutions of the RCEP, including unbounded ones. As in Chapter 3, our development will be based on convex optimization, similar in spirit to [51], for the special case of the RCEP (but readily extends to Nevanlinna-Pick interpolation case). Again, our analysis proceeds differently from [51] and continues the *partial* extension of [5] developed in the preceding chapter. We have already noted there are some important differences between the approach of that chapter and [51]. For example, results of Chapter 3, such as a certain necessary and sufficient condition for boundedness of a solution, do not follow obviously from [51]. The analysis in [51] is carried out by reposing the problem in the setting of contractive functions on the unit disc via a certain bilinear transformation. This transformation effectively avoids complications or awkward details which may arise when dealing with positive real functions. In connection with the last point, it was mentioned in the last chapter that the functional $\mathbb{K}_\Psi$ studied in [51] always has a stationary maximizer, whereas the minimizer of the functional $\mathbb{J}_\Psi$ investigated in the present and previous chapter need not be stationary. This chapter tackles the problem directly in the original positive-real setting, without recourse to the space of contractive functions. A possible advantage of this, for the special case

of the RCEP, is that the analysis is done purely on a complex Euclidean space instead of a function space as in [51]. Moreover, we show that solving the RCEP is essentially equivalent to finding the minimizers of a class of (strictly) convex functionals defined on a subset of the complex Euclidean space. This is done by establishing a new result on a bijective correspondence between denominator polynomials of non-strictly-positive solutions of the RCEP and the minimizers of the class of convex functionals associated with non-strictly-positive pseudopolynomials (Theorem 4.2.2). As a corollary to that result, we obtain an alternative and constructive derivation of Theorem 3.2.3, and a new proof of a homeomorphism which was established in [61] for the special case of real interpolators. An analogous treatment of what we accomplish here for the RCEP (and degree constrained rational interpolation in general) may also be possible in the general setting of [51] by considering some appropriate sub-class of positive real functions and establishing some additional results.

Later in Section 4.3, we generalize the homeomorphism result to also allow variation in the covariance data. In connection with this last problem, a relevant work in the literature is [62]. However, there are two features of our treatment which contrast it to [62]. The first contrasting feature is that [62] derives the unique pair of (normalized) partial covariance sequence and positive definite bounded spectral density which minimizes a certain Kullback-Leibler divergence criterion under some moment constraints, whereas here we are not interested in such an optimal pair, but we show that pairs of partial covariance sequence and pseudoplynomial data are in homeomorphic correspondence with the graph symbols of positive real rational functions of a bounded degree. In particular, we may perform a continuous coordinate transformation from the first pair to the latter pair and vice-versa. Secondly, the case where the associated pseudopolynomial is non-negative, but not positive definite, is not considered in [62]. Indeed, in this case, the solution of the RCEP may be unbounded and not integrable, while [62] restricts the solution to be integrable (see Eq. (6) therein). On the other hand, we allow for non-negative, but not positive definite, pseudopolynomials and do not impose integrability of the solutions. The importance of considering simultaneous variation of the covariance and pseudopolynomial data lies in the fact that in practice, for example in spectral estimation, both data are typically unknown and have to be estimated. Continuity implies that the resulting spectral density estimate will be robust to small errors in the estimates of the pair of data.

We also mention the paper [63] which was brought to our attention by a referee for the paper [18]. It solves a generalized moment problem with complex-

ity constraint; however, the problem treated there is rather different since the non-negative functions μ, which are monotone non-decreasing and of bounded variation on a compact interval $[a, b]$ of the real line, sought in [63] must satisfy a finite set of moment conditions and can be expressed as $\frac{d\mu}{dt} = \frac{P(t)}{Q(t)}$ for some functions $P(t)$ and $Q(t)$ which are non-negative for almost all $[a, b]$ and for which the ratio $\frac{P(t)}{Q(t)}$ is *integrable* on $[a, b]$ (the latter conditions on $\frac{d\mu}{dt}$ are also referred to collectively as "complexity constraint"). This is not the case in general for the RCEP since (unbounded) solutions f of the RCEP that have one or more poles on the unit circle *do not* correspond, via the moment constraints

$$c_0 = 2f(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta} d\mu(e^{i\theta})$$

and

$$c_k = \frac{f^{(k)}(0)}{k!} = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta} d\mu(e^{i\theta}),$$

for $k = 1, \ldots, n$, to absolutely continuous functions μ on $[-\pi, \pi]$ (see, e.g., [12, eqs. (3.10)-(3.12), p. 36]).

This chapter continues to use the notation, definitions and results of Chapter 3. However, we now also make note of the following observation. The restriction of any element of $\overline{\mathfrak{Q}_+(n, A)}\backslash\{0\}$ to $\mathbb{T}$ is a rational spectral density of McMillan degree at most $2n$, thus we shall often also view any such element as a spectral density. Hence, to each $d \in \overline{\mathfrak{Q}_+(n, A)}\backslash\{0\}$ we may associate a *unique* outer polynomial of degree at most n, denoted by $\phi(d)$, which is the unique canonical spectral (CSF) of d satisfying: $\phi(d)(0) > 0$ and $|\phi(d)(z)|^2 = d(z)\ \forall z \in \mathbb{T}$. Details on outer functions, spectral densities and CSF's can be found in [64, 15] and will also be given in the upcoming Chapter 6.

The discussion of this chapter is adapted from the paper [18].

4.2 An analysis of all solutions of the RCEP

For $\Psi \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$, consider once again the functional $\mathbb{J}_\Psi$ introduced in Chapter 3. Recall from Chapter 3 that the relationship between $\mathbb{J}_\Psi$ and the RCEP lies in its directional derivatives. For any $q, q_0 \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}$, the directional derivative at q in the direction $q_0 - q$ is defined by:

$$\nabla_{q_0 - q}\mathbb{J}_\Psi(q) = \lim_{h \downarrow 0} \frac{\mathbb{J}_\Psi(q + h(q_0 - q)) - \mathbb{J}_\Psi(q)}{h}.$$

Let $\mathcal{M}_\Psi$ denote the set of all $q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ such that all roots of $Q(q)$ on $\mathbb{T}$, including multiplicities, are all also roots of Ψ on $\mathbb{T}$. $\nabla_{q_0-q}\mathbb{J}_\Psi(q)$ is given $\forall q_0 \in \mathfrak{Q}_+(n,\mathbb{C})$ by:

$$\nabla_{q_0-q}\mathbb{J}_\Psi(q) = \begin{cases} \sum_{k=0}^{n} \Re\Big\{\Big(c_k - \langle \frac{\Psi}{Q(q)}, g_k\rangle\Big)^* (q_{0,k} - q_k)\Big\} \text{ if } q \in \mathcal{M}_\Psi \\ \infty \text{ or } -\infty \text{ otherwise} \end{cases}, \tag{4.1}$$

where $\mathrm{col}(q_0, q_1, \ldots, q_n) = q$, $\mathrm{col}(q_{0,0}, q_{0,1}, \ldots, q_{0,n}) = q_0$, and $g_k(z) = z^k$. If q_s is a stationary point, i.e., $\nabla_{q_0-q_s}\mathbb{J}_\Psi(q_s) = 0$ for all $q_0 \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}$, then $\langle \frac{\Psi}{Q(q_s)}, g_k\rangle = c_k$ for $k = 0, 1, \ldots, n$ and it follows from the Herglotz representation [65] that there is a unique $f \in \mathcal{C} \cap \mathcal{H}^\infty$ such that $f + f_* = \frac{\Psi}{Q(q_s)}$ and f is a solution of the RCEP. As we had shown in the last chapter, stationarity of q_s is in fact necessary and sufficient for a solution to be in $\mathcal{H}^\infty$ (i.e., has no poles on $\mathbb{T}$). By the strict convexity of $\mathbb{J}_\Psi$, q_s is also its unique minimizer. The point q_s could be in the interior or boundary of $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$. However, when $\Psi \in \mathfrak{Q}_+(n,\mathbb{C})$, the minimizer is guaranteed to be an interior point [5, 14]. The following lemma states this precisely. It was shown in [14] for the degree constrained Nevanlinna-Pick interpolation problem, but which by inspection holds analogously for the RCEP (actually, [14] shows that both a and b have no roots on $\mathbb{D}$, that the same is true for $a + b$ follows from [9, Proposition 2.6] or by simply noting that $\frac{a}{b} + 1 \in \mathcal{C}_+$).

Lemma 4.2.1 *If $\Psi \in \mathfrak{Q}_+(n,\mathbb{C})$ then $q_{\min}$, the unique minimizer of $\mathbb{J}_\Psi$, is a stationary point in $Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$. Furthermore, if b is the CSF of $Q(q_{\min})$ and a is determined uniquely from $a_*b + ab_* = \Psi$, then (a, b) is the unique pair with $b(0) > 0$ such that $a_*b + ba_* = \Psi$, $a + b$ has no roots on $\mathbb{D}$ and $f = \frac{a}{b} \in \mathcal{C}_+$ is a solution of the RCEP.*

We shall use Theorem 3.3.8 and Lemma 4.2.1 as the basis of our analysis of the RCEP. To this end, we introduce the following notation: If $p(z) = \sum_{k=0}^{n} p_k z^k$ is a polynomial of degree at most n then α_p is defined as $\mathrm{col}(p_0, p_1, \ldots, p_n)$. Hence, p can be written as $p(z) = \alpha_p^{\mathrm{T}} Z_n(z)$, where Z_n is a complex vector-valued function defined by $Z_n(z) = \mathrm{col}(1, z, \ldots, z^n)$. We also define $\|g\|_\infty = \operatorname*{ess\,sup}_{z\in\mathbb{T}} |g(z)|$ and $\|g\|_2 = (\langle g, g\rangle)^{\frac{1}{2}}$ for any complex function g which is measurable on $\mathbb{T}$. In the ensuing analysis, we shall make use of the following observation. By an argument given in [61, Appendix A], the interpolation constraints

$$f(0) = \frac{1}{2}c_0,\ f^{(k)}(0) = c_k \text{ for } k = 1, \ldots, n, \tag{4.2}$$

imply that there is a $(n+1)\times(n+1)$ matrix W, whose entries are continuously dependent on the value of the associated PCS $c_0, c_1, \ldots, c_n$, such that if a, b are polynomials of degree at most n with $b(0) > 0$, and $f = \frac{a}{b}$ satisfies (4.2), then α_a and α_b are linearly related via $\alpha_a = W\alpha_b$ (note that f need not be Carathéodory). To emphasize the (continuous) dependence of W on $c = (c_0, c_1, \ldots, c_n)$, we shall at times write $W(c)$ in place of W. The next theorem, which we are now in a position to show, extends the known Lemma 4.2.1 to $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$:

Theorem 4.2.2 *Let Ψ be an arbitrary element of $\partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$ and let $q_{\min}$ be as in Theorem 3.3.8. If b is the CSF of $Q(q_{\min})$ and a is such that $\alpha_a = W\alpha_b$, then (a, b) is the unique pair such that $b(0) > 0$, $a+b$ has no roots on $\mathbb{D}$, $a_*b + ab_* = \Psi$ and $f = \frac{a}{b}$ is a solution of the RCEP.*

Proof. Let the sequences $\{\Psi_k\}_{k\geq 1}$ and $\{Q(q_{\min,k})\}_{k\geq 1}$ be as in Theorem 3.3.8. Then $\{Q(q_{\min,k})\}_{k\geq 1}$ is a sequence of rational spectral densities having the properties:

i) there can be at most $2n$ roots of $\{Q(q_{\min,k})\}_{k\geq 1}$ on or approaching $\mathbb{T}$ as $k \to \infty$, and

ii) it is uniformly bounded in magnitude by some positive number M (by Theorem 3.3.8).

It follows that the sequence $\{Q(q_{\min,k})\}_{k\geq 1}$ of spectral densities satisfy a set of sufficient conditions given in [66, Theorem 8] (Theorem 6.4.5 of Chapter 6) which guarantee $\{\log Q(q_{\min,k})\}_{k\geq 1}$ to be *uniformly integrable* on $\mathbb{T}$, i.e.,

$$\lim_{c\to\infty} \sup_{k\geq 1} \langle I_{\{|\log Q(q_{\min,k})(e^{i\theta})|>c\}}, |\log Q(q_{\min,k})|\rangle = 0,$$

where I_A is the indicator function for the set A. Let $b_k = \phi(Q(q_{\min,k}))$ and $b = \phi(Q(q_{\min}))$. Since $Q(q_{\min,k}) \overset{\|\cdot\|_\infty}{\to} Q(q_{\min})$ (Theorem 3.3.8) and $\{\log Q(q_{\min,k})\}_{k\geq 1}$ is uniformly integrable, it follows from [15] that $b_k \overset{\|\cdot\|_2}{\to} b$. However, since $b_k - b$ is a polynomial of degree at most n for all k, we also have that $b_k \overset{\|\cdot\|_\infty}{\to} b$. Before proceeding further, we make the following observation. From Lemma 4.2.1 we note that when $\Psi_k \in \mathfrak{Q}_+(n, \mathbb{C})$ then $a_k + b_k$ has no roots on $\mathbb{T}$ and $f_k = \frac{a_k}{b_k}$ is a solution of the RCEP if a_k is uniquely determined from the equation $a_{k*}b_k + a_k b_{k*} = \Psi_k$. However, as given in [14, p. 831], the last equation is equivalent to solving the linear equation $S(\alpha_{b_k})\alpha_{a_k} = d_k$ for α_{a_k} with $d_k = Q^{-1}(\Psi_k)$ and S being a continuous linear operator from $\mathbb{R}\times\mathbb{C}^n$ to $\mathbb{C}^{n+1}\times\mathbb{C}^{n+1}$ (actually, [14]

considers the case where b and $a+b$ are free of roots on $\mathbb{C}\backslash\overline{\mathbb{D}}$, but an analogous argument holds in our setting). Uniqueness of α_{a_k} (hence also of a_k) follows from non-singularity of $S(\alpha_{b_k})$. However, given b_k, we also know that a_k must satisfy $\alpha_{a_k} = W\alpha_{b_k}$. Therefore, when $\Psi_k \in \mathfrak{Q}_+(n,\mathbb{C})$, determining a_k by solving $a_k b_{k*} + a_{k*} b_k = \Psi_k$ (as stated in Lemma 4.2.1) or via the relation $\alpha_{a_k} = W\alpha_{b_k}$ are *equivalent*. Continuing on with our proof, let $a_k = \alpha_{a_k}^{\mathrm{T}} Z_n$ with $\alpha_{a_k} = W\alpha_{b_k}$. Since $\|b_k - b\|_\infty \to 0$ as $k \to \infty$ it follows that $||W\alpha_{b_k} - W\alpha_b||_2 \to 0$ as $k \to \infty$. Defining a via $\alpha_a = \lim_{k\to\infty} W\alpha_{b_k} = W\alpha_b$, we conclude that $f = \frac{a}{b}$ satisfies (4.2). For the remaining parts of the proof, we may assume Ψ_k has no roots on $\mathbb{T}$ for all k. There is no loss in generality in taking this assumption since $q_{\min}$, as the limit of $q_{\min,k}$, is independent of the particular sequence $\{\Psi_k\}_{k\geq 1}$ used in approaching Ψ. Then we have from Lemma 4.2.1 that $f_k = \frac{a_k}{b_k}$ is a solution of the RCEP for all k. All that remains now is to show that $f \in \mathcal{C}$. Define the set $r(b) = \{z \in \mathbb{T} \mid b(z) = 0\}$. Since $b_k \to b$ and $a_k \to a$ uniformly on $\overline{\mathbb{D}}$ (b_k and a_k are in $\mathcal{H}^\infty$ and continuous for all k), and b_k and a_k have no roots in $\overline{\mathbb{D}}$ for all k (due to Lemma 4.2.1 and our assumption of positive definiteness of Ψ_k), we obtain:

$$\Re\Big\{\frac{a(z)}{b(z)}\Big\} = \Re\Big\{\lim_{k\to\infty}\frac{a_k(z)}{b_k(z)}\Big\} = \lim_{k\to\infty}\Re\Big\{\frac{a_k(z)}{b_k(z)}\Big\} \geq 0 \;\forall z \in \overline{\mathbb{D}}\backslash r(b).$$

Let us now consider points z in $r(b)$. To this end, let $z_0 \in r(b)$ be such that z_0 is also a root of a. If z_0 has the same multiplicity as a root of a as it does as a root of b then there is cancellation between the polynomials a and b, f is continuous at z_0, and it follows that $\Re\{f(z_0)\} = \lim_{z\to z_0}\Re\big\{\frac{a(z)}{b(z)}\big\} \geq 0$ since $\Re\{f\} \geq 0$ on $\overline{\mathbb{D}}\backslash r(b)$. For all other $z \in r(b)$, it is straightforward to see, again since $\Re\{f\} \geq 0$ on $\overline{\mathbb{D}}\backslash r(b)$, that $\Re\big\{\frac{a(z)}{b(z)}\big\} = \infty$. Thus we conclude $\Re\big\{f(z)\big\} \geq 0$ for $\forall z \in \overline{\mathbb{D}}$, i.e., $f \in \mathcal{C}$. Finally, since $a_k b_{k*} + a_{k*} b_k = \Psi_k$, and $a_k + b_k$ has no roots on $\mathbb{D}$, for each k, by taking passage to the limit as $k \to \infty$ we easily see that a and b must satisfy $a_* b + a b_* = \Psi$ and $a + b$ also has no roots on $\mathbb{D}$.

We shall now show the converse: If (a,b) is any pair with $b(0) > 0$ such that $a+b$ has no roots on $\mathbb{D}$, $f = \frac{a}{b}$ is a solution of the RCEP (hence $\alpha_{a_k} = W\alpha_{b_k}$ is automatically satisfied) and $ab_* + ba_* = \Psi$, then necessarily $b = \phi(Q(q_{\min}))$. If $f \in \mathcal{H}^\infty \cap \mathcal{C}$ then we have $\langle f + f_*, g_k\rangle = c_k$ for $k = 0, 1, \ldots, n$ (recall that $g_k = z^k$). Noting that $f + f_* = \frac{a_* b + b a_*}{b b_*}$, from (4.1) we see that $Q^{-1}(b b_*)$ coincides with $q_{\min}$, the unique minimizer of $\mathbb{J}_\Psi$ with $\Psi = a_* b + a b_*$. Therefore, $b = \phi(Q(q_{\min}))$. Suppose now that f has one or more poles on $\mathbb{T}$. Note that such poles can only be simple (i.e., poles of multiplicity 1) [12, p. 35-36]. Let us write $f = \frac{a}{b_1 b_0}$, where $b_1(0)b_0(0) > 0$, $\frac{a}{b_1} \in \mathcal{H}^\infty$, b_0 has all its zeros on $\mathbb{T}$, and b_0 and $\frac{a}{\gcd(a,b_1)}$ are co-prime

($\gcd(a, b_1)$ denotes the unique monic polynomial which is the greatest common divisor of a and b_1). Let $b_0 = \Pi_{k=1}^{m}(1 - e^{-i\theta_k}z)$ with $m < n$ and $\theta_1, \ldots, \theta_m \in (-\pi, \pi]$ ($\theta_i \neq \theta_j$ whenever $i \neq j$). Now, we may decompose f as (see, e.g., [12, eqs. (3.11)-(3.12)]):

$$f = \frac{a_1}{b_1} + \sum_{l=1}^{m} \frac{K_l}{2} \frac{1 + e^{-i\theta_l}z}{1 - e^{-i\theta_l}z},$$

where $\frac{a_1}{b_1} = f_1 \in \mathcal{H}^\infty \cap \mathcal{C}$ and K_l are positive constants for $l = 1, \ldots, m$. Note that for each l, the term $\frac{K_l}{2}\frac{1+e^{-i\theta_l}z}{1-e^{-i\theta_l}z}$ is in $\mathcal{C}$ and has a pole at $z = e^{i\theta_l}$. Moreover, if $\Psi_1 = a_1 b_{1*} + a_{1*} b_1$ then, by the argument directly above [12, eq. (3.13)], we also have that $\Psi = a_* b + a b_* = \Psi_1 b_{0*} b_0$. Let us now define $f_2 \in \mathcal{C}$ by

$$\begin{aligned} f_2(z) &= f(z) - f_1(z) \\ &= \sum_{l=1}^{m} \frac{K_l}{2} \frac{1 + e^{-i\theta_l}z}{1 - e^{-i\theta_l}z} \\ &= \sum_{l=1}^{m} K_l \Big(\frac{1}{2} + \sum_{k=1}^{\infty} e^{-ik\theta_l} z^k\Big), \text{ for all } z \in \mathbb{D}. \end{aligned}$$

Clearly, we may also write $f_2(z) = \frac{1}{2}c_{2,0} + \sum_{k=1}^{n} c_{2,k} z^k +$ higher order terms, where $c_{2,0} = \sum_{l=1}^{m} K_l$ and $c_{2,k} = \sum_{l=1}^{m} K_l e^{-ik\theta_l}$ for $k = 1, \ldots, n$. Since $f_1 \in \mathcal{H}^\infty \cap \mathcal{C}$ and $f_1 + f_{1*} = \frac{\Psi_1}{b_1 b_{1*}} = \frac{\Psi}{b_0 b_1 b_{0*} b_{1*}}$, from the definition $f_2 = f - f_1$ it follows that $c_{2,0}, c_{2,1}, \ldots, c_{2,n}$ are also given by $c_{2,k} = c_k - \langle f_1 + f_{1*}, g_k\rangle = c_k - \langle \frac{\Psi}{b_1 b_0 b_{1*} b_{0*}}, g_k\rangle$ for $k = 0, \ldots, n$. Defining $q_1 = \operatorname{col}(q_{1,0}, q_{1,1}, \ldots, q_{1,n}) = Q^{-1}(b_1 b_0 b_{1*} b_{0*})$, we have for any $q = \operatorname{col}(q_0, q_1, \ldots, q_n) \in \overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))}$:

$$\begin{aligned} \nabla_{q-q_1} \mathbb{J}_\Psi(q_1) &= \Re\Big\{ \sum_{k=0}^{n} \Big(c_k - \langle \frac{\Psi}{Q(q_1)}, g_k\rangle \Big)^* (q - q_1) \Big\} \\ &= \Re\big\{ \sum_{l=1}^{m} K_l (q_0 - q_{1,0}) + \sum_{k=1}^{n} \Big(\sum_{l=1}^{m} K_l e^{-ik\theta_l} \Big)^* (q_k - q_{1,k}) \big\} \\ &= \sum_{l=1}^{m} K_l \Re\big\{ (q_0 - q_{1,0}) + \sum_{k=1}^{n} e^{ik\theta_l} (q_k - q_{1,k}) \big\} \\ &= \sum_{l=1}^{m} K_l Q(q - q_1)(e^{i\theta_l}). \end{aligned}$$

Now, since $Q(q_1)(e^{i\theta_l}) = 0$ for $l = 1, \ldots, m$, it is clear that $Q(q - q_1)(e^{i\theta_l}) = Q(q)(e^{i\theta_l}) \geq 0$ for $l = 1, \ldots, m$. We conclude that $\nabla_{q-q_1} \mathbb{J}_\Psi(q_1) \geq 0 \ \forall q \in \overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))}$. By strict convexity of $\mathbb{J}_\Psi$ it follows that q_1 coincides with $q_{\min}$,

the unique minimizer of $\mathbb{J}_\Psi$. Since $b_1(0)b_0(0) > 0$ and recalling the definition of q_1, this proves that $b_1 b_0 = \phi(Q(q_{\min}))$. □

The above theorem is a new result. The main idea of the proof is to show that the bijective mapping from $\Psi \in \mathfrak{Q}_+(n, \mathbb{C})$ to (a, b) in Lemma 4.2.1 continues to hold for $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$. It is interesting because it reveals that, for any $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$, $\phi(Q(q_{\min}))$ is actually also a unique denominator polynomial of some solution of the RCEP. Combining this with Theorems 3.3.3 and 3.3.7 of Chapter 3, we see there are three possible scenarios when $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$:

1. $q_{min} \in \mathfrak{Q}_+(n, \mathbb{C})$ and is a stationary point of $\mathbb{J}_\Psi$
2. $q_{min} \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$ and is a stationary point of $\mathbb{J}_\Psi$
3. $q_{min} \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$ but is *not* a stationary point of $\mathbb{J}_\Psi$. This is precisely the case when the associated interpolant has a pole on the unit circle.

The three possible scenarios are illustrated in Fig. 4.1.

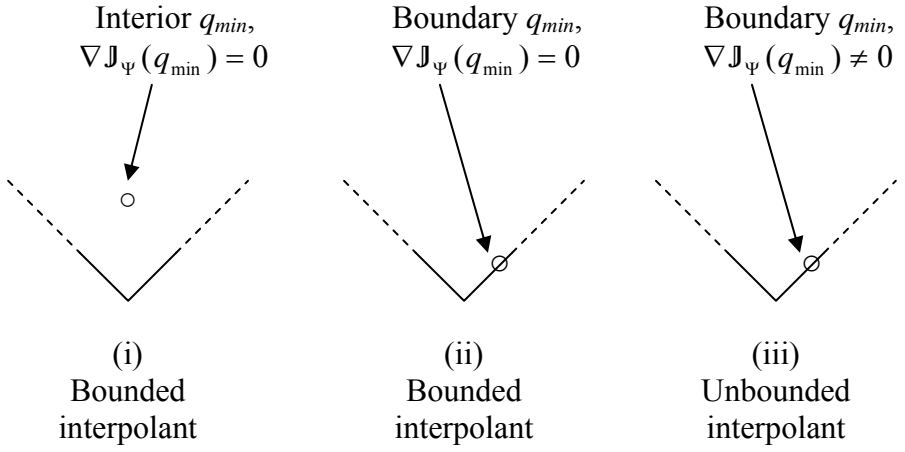

Figure 4.1: Three possible scenarios when $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$

Now, combining Theorem 4.2.2 with Lemma 4.2.1 we have the following corollary:

Corollary 4.2.3 *Let Ψ be an arbitrary element of $\overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$ and let $q_{\min}$ be as in Theorem 3.3.8. If b is the CSF of $Q(q_{\min})$ and a is such that $\alpha_a = W\alpha_b$, then (a, b) is the unique pair such that $b(0) > 0$, $a+b$ has no roots on $\mathbb{D}$, $a_* b + a b_* = \Psi$ and $f = \frac{a}{b}$ is a solution of the RCEP.*

Therefore, to every functional in the set $\{\mathbb{J}_\Psi\}_{\Psi\in\overline{\mathfrak{Q}_+(n,\mathbb{C})}\setminus\{0\}}$, there is exactly one pair (a,b) with $b(0)>0$ and $a+b$ having no roots on $\mathbb{D}$ such that $f=\frac{a}{b}$ is a solution of the RCEP, and vice-versa. As a byproduct we obtain a new constructive proof, via the set of functionals $\{\mathbb{J}_\Psi\}_{\Psi\in\overline{\mathfrak{Q}_+(n,\mathbb{C})}\setminus\{0\}}$, of Theorem 3.2.3 on complete parametrization of all solutions of the RCEP by elements of $\overline{\mathfrak{Q}_+(n,\mathbb{C})}\setminus\{0\}$. As another byproduct of Corollary 4.2.3, we also obtain a homeomorphism which was first shown in [61] for real solutions of the RCEP.

Corollary 4.2.4 *Let* $\mathcal{D}=\left\{d\in\mathbb{R}\times\mathbb{C}^n \mid Q(d)\in\overline{\mathfrak{Q}_+(n,\mathbb{C})}\setminus\{0\}\right\}$ *and* $\mathcal{A}=\left\{b\in(0,\infty)\times\mathbb{C}^n \mid b^{\mathrm{T}}Z_n \text{ is outer and } \frac{(Wb)^{\mathrm{T}}Z_n}{b^{\mathrm{T}}Z_n} \text{ is a solution of the RCEP}\right\}$. *Then the map* $\mathrm{G}:\mathcal{D}\to\mathcal{A}$ *defined by* $\mathrm{G}:d\mapsto\alpha_{g(d)}$ *is a homeomorphism, where* $g(d)=\phi(Q(q_d))$, $q_d=\underset{q\in\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}}{\arg\min}\mathbb{J}_{\Psi_d}(q)$, *and* $\Psi_d(z)=d_0+\sum_{k=1}^{n}\left(d_kz^k+d_k^*z^{-k}\right)$.

Proof. That G is a bijection is already clear from results preceding the corollary. Thus we only have to show that G and G^{-1} are continuous. We will do this for G, the same follows for G^{-1} by a similar argument. Let d be any element in $\mathcal{D}$ and define $\mathcal{B}_d$ to be the set of all infinite sequences $d_1,d_2,\ldots$ of elements in $\mathcal{D}$ such that $\|d_k-d\|_2\overset{k\to\infty}{\to}0$. In first paragraph of the proof of Theorem 4.2.2 we have shown that $\|g(d_k)-g(d)\|_\infty\overset{k\to\infty}{\to}0$ for *all* sequences in $\mathcal{B}_d$. Hence also $\|\alpha_{g(d_k)}-\alpha_{g(d)}\|_2\overset{k\to\infty}{\to}0$ for all such sequences. Suppose that G is not continuous at d, then there must exist a sequence $e_1,e_2,\ldots$ of elements in $\mathcal{D}$ such that $\|e_k-d\|_2\overset{k\to\infty}{\to}0$ but for which $\|\alpha_{e_k}-\alpha_d\|_2\overset{k\to\infty}{\nrightarrow}0$. This is a contradiction, so G must be a continuous map and the proof is complete. □

4.3 Generalization of results to simultaneous variation of covariance and pseudopolynomial data

Thus far we have only looked at the continuous relationship between Ψ and b when Ψ is varied and the PCS is fixed. However, the ideas used in deriving Theorem 3.3.8, Theorem 4.2.2 and Corollary 4.2.4 can be adapted easily to analyze the case where the PCS is allowed to vary. In this section we shall state generalizations of Theorem 3.3.8 and Corollary 4.2.4. Since the main ideas here are the same as in the last section, we shall only sketch the proofs.

Let $\mathcal{P}=\{(c_0,c_1,\ldots,c_n)\in\mathbb{R}\times\mathbb{C}^n \mid c_0,c_1,\ldots,c_n \text{ is a PCS of order } n\}$ and define the functional $\mathbb{M}_{c,\Psi}$ exactly as on the right hand side of (3.3), but we now

consider c to be an additional parameter of the function alongside Ψ. Notice that $\mathcal{P}$ is a convex set. Then we have the following analogue of Theorem 3.3.8:

Lemma 4.3.1 *Let $\Psi, \Psi_k, q_{\min}, q_{\min,k}$ be as defined in Theorem 3.3.8 with $\mathbb{J}_{\Psi_k}$ and $\mathbb{J}_{\Psi}$ replaced by $\mathbb{M}_{c_k,\Psi_k}$ and $\mathbb{M}_{c,\Psi}$, respectively, with $c_k, c \in \mathcal{P}$. If $\lim_{k\to\infty} \|c_k - c\|_2 = 0$ then $\lim_{k\to\infty} \|q_{\min} - q_{\min,k}\|_2 = 0$ and $\lim_{k\to\infty} \|Q(q_{\min}) - Q(q_{\min,k})\|_\infty = 0$.*

Proof. As in the proof of Theorem 3.3.8, let $X_\epsilon(q_{\min}) = \{q \in \mathbb{R} \times \mathbb{C}^n \mid \|q - q_{\min}\| \leq \epsilon\} \cap \overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))}$ with $\epsilon > 0$ small enough such that $0 \notin X_\epsilon(q_{\min})$. Then $|\mathbb{M}_{c,\Psi}(q) - \mathbb{M}_{c_k,\Psi_k}(q)| \leq \|c - c_k\|_2 D_1 + \|\Psi - \Psi_k\|_\infty D_2$ for all $q \in X_\epsilon(q_{\min})$, where $D_1 = \max_{X_\epsilon(q_{\min})} \|q\|_2$, $D_2 = \max_{X_\epsilon(q_{\min})} \langle \mathbf{1}, |\log Q(q)| \rangle$ and $\mathbf{1} : z \mapsto 1 \, \forall z \in \mathbb{T}$. The remainder of the proof proceeds along similar lines to the proof of Theorem 3.3.8 by taking a suitably large k so that both $\|c - c_k\|_2$ and $\|\Psi - \Psi_k\|_\infty$ are sufficiently small. □

Theorem 4.3.2 *Let $\mathcal{D}$ be as in Corollary 4.2.4 and define $\mathcal{S} = \{(a, b) \in \mathbb{C}^{n+1} \times (0, \infty) \times \mathbb{C}^n \mid b^{\mathrm{T}} Z_n$ is outer and $f = \frac{a^{\mathrm{T}} Z_n}{b^{\mathrm{T}} Z_n} \in \mathcal{C}\}$. Then the map $\mathrm{H} : \mathcal{P} \times \mathcal{D} \to \mathcal{S}$ defined by $\mathrm{H} : (c, d) \mapsto (W(c)\alpha_{g(c,d)}, \alpha_{g(c,d)})$ is a homeomorphism, where $g(c, d) = \phi(Q(q_{(c,d)}))$, $q_{(c,d)} = \operatorname{arg\,min}_{q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}} \mathbb{M}_{c,\Psi_d}(q)$, and $\Psi_d(z)$ is as defined in Corollary 4.2.4.*

Remark 4.3.3 *Note that, as stated in the Introduction, a stronger version of the above theorem for the domain $\mathcal{P} \times$ Interior of $\mathcal{D}$ has been given, by differential geometric techniques, in [67, Theorem 6.6] in the context of the Nevanlinna-Pick interpolation problem with degree constraint. Our innovation in the theorem is extending the weaker property of homeomorphism to $\mathcal{P} \times \partial\mathcal{D}$ by utilizing a different technique.*

Proof. That H is surjective (onto) follows from the definition. Injectivity of H is also easily established since each pair (a, b) uniquely defines c by the first $n + 1$ coefficients of the Taylor series expansion of $f = \frac{a}{b}$ about $z = 0$ and d via the relation $\Psi_d = a^{\mathrm{T}} Z_n (b^{\mathrm{T}} Z_n)_* + (a^{\mathrm{T}} Z_n)_* (b^{\mathrm{T}} Z_n)$. Thus, H is a bijective map.

We now show that H is continuous. Let (c, d) be an arbitrary element of $\mathcal{P} \times \mathcal{D}$ and $(a, b) = H(c, d)$. Let $\{d_k\}_{k\geq 1}$ be any sequence in $\mathcal{D}$ such that $\|d_k - d\|_2 \overset{k\to\infty}{\to} 0$. We also let $\{c_k\}_{k\geq 1}$ be a sequence in $\mathcal{P}$ such that $\|c_k - c\|_2 \overset{k\to\infty}{\to} 0$. If $(a_k, b_k) = H((c_k, d_k))$ then from Lemma 4.3.1 we have that $\|b - b_k\|_2 \overset{k\to\infty}{\to} 0$. Recall that $a_k = W(c_k) b_k$ for $k = 1, 2, \ldots$ and $a = W(c)b$. Since $\|(c_k, d_k) - (c, d)\|_2 \overset{k\to\infty}{\to} 0$, we have that $\|W(c) - W(c_k)\|_\infty \overset{k\to\infty}{\to} 0$ (here $\|\cdot\|_\infty$ denotes the operator-norm of a

matrix) and necessarily $\|a_k - a\|_2 = \|W(c_k)b_k - W(c)b\|_2 \overset{k\to\infty}{\to} 0$. The continuity of H then follows from arguments similar to those in the final part of the proof of Corollary 4.2.4.

The remaining part of the proof is only to show that H^{-1} is continuous. This is also quite straightforward. Let $(c_i, d_i) = H^{-1}((a_i, b_i))$ for $i = 1, 2$. If $\|(a_1, b_1) - (a_2, b_2)\|_2$ is small enough it follows that $\|a_1^{\mathrm{T}} Z_n - a_2^{\mathrm{T}} Z_n\|_\infty$ and $\|b_1^{\mathrm{T}} Z_n - b_2^{\mathrm{T}} Z_n\|_\infty$ will also be small. Letting $f_1 = \frac{a_1^{\mathrm{T}} Z_n}{b_1^{\mathrm{T}} Z_n}$ and $f_2 = \frac{a_2^{\mathrm{T}} Z_n}{b_2^{\mathrm{T}} Z_n}$, this implies that $|f_1(z) - f_2(z)|$ will be uniformly small on any *compact subset* K of $\mathbb{D}$. From this we see that necessarily $\|c_1 - c_2\|_2$ is small by examination of the Cauchy-integrals $|\oint_C \frac{f_1(z) - f_2(z)}{z^{l+1}} dz|$ for $l = 0, 1, \ldots, n$ over some closed path C in K which encloses the point $z = 0$. Since $\|d_1 - d_2\|_2$ is trivially small when $\|(a_1, b_1) - (a_2, b_2)\|_2$ is small, we conclude that $\|(c_1, d_1) - (c_2, d_2)\|_2$ is small. Although the preceding argument is purely formal, it can easily be made rigorous and we may conclude that H^{-1} is continuous. This completes the proof. □

By Theorem 4.3.2 we may perform a continuous coordinate transformation from (a, b) to (c, d) and vice-versa. The theorem may also be interpreted as saying that graph symbols of all positive real functions of degree at most n are parametrized by pairs of PCS of order n and non-negative pseudopolynomial data. There are also previous results which deal with variation in c with d fixed, a, b, c, d real, and Ψ_d not having roots on $\mathbb{T}$, and again under those special assumptions a stronger result of diffeomorphism can be shown [68, 6]. A work which considers uncertain covariance data is [62]. However, as noted in the Introduction, the purpose and results of that work are rather different to ours.

4.4 Concluding remarks

This chapter gives an analysis of the RCEP which yields new results in Theorem 4.2.2 for solutions parametrized by pseudopolynomials in $\partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$ and in the part of Theorem 4.3.2 which extends the domain of the homeomorphism to $\mathcal{P} \times \partial\mathcal{D}$.

Previously, it has been shown that any real solution of the RCEP can be found by solving non-linear equations for α_b [61]. Corollary 4.2.3 of this paper shows that elements of α_b are actually the coefficients of $\phi(Q(q_{\min}))$, where $q_{\min}$ is the minimizer of $\mathbb{J}_\Psi$. We are led to the following conclusion:

Solving the RCEP is essentially equivalent to
finding the minimizer of $\mathbb{J}_\Psi$ for all $\Psi \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$.

From Corollary 4.2.3 also follow a new proof of a theorem by Georgiou on complete parametrization of all solutions of the RCEP and a new proof, with an extension to non-real interpolators, of a homeomorphism which was established in [49]. Theorem 4.3.2 is a generalization of this homeomorphism.

Differences between our convex optimization treatment and the extensive and abstract generalization of [5] given in [51], some of which have been discussed in the Section 3.4, can now be seen more clearly. The most of important of these is that in [51], the case $\Psi \in \partial\mathfrak{Q}_+(n,\mathbb{C})\backslash\{0\}$ is treated via analysis of a functional $\mathbb{K}_\Psi$ defined on a set of Schur functions satisfying a certain constraint (refer to the discussion in Section 3.4). This amounts to solving an optimization problem over a *function space*; a solution f of the RCEP is obtained directly as the unique maximizer of $\mathbb{K}_\Psi$. Our treatment is via analysis of a functional $\mathbb{J}_\Psi$ defined on a convex subset of $\mathbb{R} \times \mathbb{C}^n$ (which is a *finite dimensional space*) and by allowing Ψ to have zeros on $\mathbb{T}$. From $\mathbb{J}_\Psi$ we do not directly obtain a solution f of the RCEP, but a unique denominator polynomial of f. The associated numerator polynomial is then determined by the PCS and the denominator polynomial. A possible advantage of $\mathbb{J}_\Psi$, since it is defined on a closed, convex subset of $\mathbb{R} \times \mathbb{C}^n$, is that it could be amenable to numerical optimization algorithms. On the other hand, [51] does not investigate how to (numerically) compute the unique maximizer of $\mathbb{K}_\Psi$ in the space of Schur functions, and the discussion of the RCEP in Section 5.1 therein is limited to the case where $\Psi \in \mathfrak{Q}_+(n,\mathbb{C})$.

Although the functional $\mathbb{K}_\Psi$ always has a stationary maximizer, the difficulty with numerical optimization of $\mathbb{K}_\Psi$ is that it is defined on a function space of contractive functions. Implementation of an algorithm for finding the maximizer of $\mathbb{K}_\Psi$ on a finite memory computer would be impossible unless the trajectory of the algorithm can be guaranteed to remain in a set of finitely-parametrized functions. An alternative solution might be to consider the dual of $\mathbb{K}_\Psi$ which may be definable on a subset of some finite-dimensional space. On the other hand, a pitfall of numerical optimization of the functional $\mathbb{J}_\Psi$ is that its minimizer need not be stationary (thus it may be difficult to "find" it) and the condition number of the Hessian explodes for points tending to the boundary. However, it turns out that these are problems which can be circumvented. As we shall see in the following chapter, the optimization problem can always be reformulated to one in which the minimizer is *always stationary*, regardless of whether Ψ is strictly positive or merely non-strictly positive, and in which the new functional to be minimized is better behaved around and on the boundary.

For further research, we may pose the question of whether Theorems 4.2.2 and

4.3.2 can be exploited to develop new fast and reliable numerical algorithms for computing the minimizer of $\mathbb{J}_\Psi$, especially for the more challenging cases where it is not a stationary point and/or lies very close to or on the boundary. Naturally, efficient/fast computation of solutions is an important practical issue. As a step in this direction, in Chapter 5 we analyze the homotopy continuation algorithm due to Enqvist [50] for the case where Ψ is allowed to be non-strictly positive.

Note that it may be possible to give a treatment analogous to ours in the general setting of [51] by considering some suitable sub-class of positive real-functions. It may also be possible that the framework developed here can be generalized to the case of matrix-valued RCEP. This can be considered in future research. Moreover, as with the results of Chapter 3, the results of this chapter also carry over to the setting of Nevanlinna-Pick interpolation with degree constraint and will now be taken up in the next chapter.

Chapter 5

Computation of Degree Constrained Rational Interpolants

5.1 Background and motivation

In Chapter 3 we had derived a characterization of bounded solutions of the RCEP which was followed up in Chapter 4 with some new results pertaining to all solutions of the RCEP, bounded or unbounded. The present chapter deals with computation of solutions of the RCEP, by exploiting some of the results of Chapter 4. However, we now derive our results in the more general setting of Nevanlinna-Pick interpolation with derivative constraints (i.e., involving interpolation constraints of the form $\frac{1}{k!}f^{(k)}(z) = w$ for some positive integer k), of which the rational covariance extension problem may be viewed as a special case. We had already mentioned that the analysis of the preceding two chapters carry over mutatis mutandis to this more general setting (which will be formalized shortly). The required adaptations will become more clear as we proceed through this chapter. To this end, let there be given $\{z_0, z_1, \ldots, z_n\} \subset \mathbb{D}$ and $\{w_0, w_1, \ldots, w_n\} \subset \{z \in \mathbb{C} \mid \Re\{z\} \geq 0\}$. We make the convention that non-unique z_k's are ordered sequentially. Moreover, for simplicity we shall assume $z_0 = 0$ and w_0 is real. There is no loss in generality in taking this assumption since the map $z \mapsto \frac{z-z_0}{1-z_0^* z}$ sends any $z_0 \in \mathbb{D}$ to 0 and is a bianalytic map from $\mathbb{D}$ onto itself. Secondly, we are allowed to subtract the imaginary part of w_0 from $w_1, \ldots, w_n$ without changing Problem 5.1.1 to be stated below. For further details, the reader may consult [60, Appendix A]. We consider the following degree

constrained rational interpolation problem:

Problem 5.1.1 *Find all $f \in \mathcal{C}$ of McMillan degree at most n such that, for $k = 0, 1, \ldots, n$, $f(z_k) = w_k$ if z_k is of multiplicity 1, and $\frac{1}{j!}f^{(j)}(z_k) = w_k$ if z_k is of multiplicity $m > 1$ and $z_k = z_{k+1} = \ldots = z_{k+m-1}$.*

As with the RCEP, it is well known that the above problem has a solution if and only if a certain *(generalized) Pick matrix*, constructed from the data $\{w_0, w_1, \ldots, w_n\}$, is non-negative definite [48, 69]. In the case of the RCEP this Pick matrix corresponds to the Toeplitz matrix T in Definition 3.2.1. Also as before, the solution is unique if the matrix is singular, otherwise there are infinitely many solutions. The following generalized version of Theorem 3.2.3 holds for Problem 5.1.1:

Theorem 5.1.2 ([48, 51, 18]) *For a given interpolation data with a positive definite Pick matrix, and any monic polynomial $\eta \neq 0$ of degree n with roots in $\overline{\mathbb{D}}$, there exists a unique pair of polynomials (a, b) of degree $\leq n$ such that $b(0) > 0$, $a + b$ has all its roots in $\mathbb{C}\backslash\mathbb{D}$, the pair satisfies the relation*

$$ab_* + ba_* = \kappa^2 \eta\eta_* \tag{5.1}$$

for a fixed $\kappa > 0$, and $f = \frac{a}{b}$ is a solution of Problem 5.1.1.

Again, roots of the polynomial η in the theorem are referred to in the literature as "spectral zeros." Problem 5.1.1 is of significance since there are many engineering problems which can be reformulated into an interpolation problem, whilst the degree constraint is naturally desirable from a practical point of view as lower degree solutions typically mean simpler controllers, filters, etc. Some of these applications include high resolution spectral estimation [70, 14], maximal power transfer [14], and robust control [60, 55, 71, 58].

A convex optimization approach for solving Problem 5.1.1, which parallels the approach we have already seen for the RCEP in Chapters 3 and 4, was initially developed in the papers [5, 14] for real η with roots inside the unit circle. However, this method, without modification, suffers the same shortcoming as remarked in Remark 3.3.6, i.e., it has features which makes it numerically unsuitable for computation of solutions with poles close to or on the unit circle. A modification of the method, by reparametrization and application of a homotopy continuation method, was first introduced by Enqvist [50] for the rational covariance extension problem, and subsequently adapted by Nagamune [72, 59], and Blomqvist

and Nagamune [60, 58] to Nevanlinna-Pick interpolation and moment problems. However, the approach had not been studied and extended to the case where η has roots on the unit circle. From a practical point of view, spectral zeros on the unit circle are important because they are associated with solutions of Problem 5.1.1 with poles on the unit circle (which, for instance, correspond to spectral lines in spectral estimation [69]) and with solutions for which the restriction of $f + f_*$ to $\mathbb{T}$ is an absorption spectrum (i.e., a spectrum with some frequencies having "zero spectral energy"), and, in control applications, solutions with poles close to $\mathbb{T}$ are often required, making it essential to have a reliable method for computing them. In [72, 59], it was demonstrated that the homotopy continuation method appears to be numerically robust and can compute solutions with poles very close to $\mathbb{T}$, but it has never been clear how close the poles can be to $\mathbb{T}$ for the method to still perform satisfactorily. This chapter shows that it can in fact compute solutions with poles *anywhere* on $\mathbb{C}\backslash\mathbb{D}$.

No "complete" algorithm has been presented for the case where some spectral zeros are on the unit circle, apart from [49]. The latter algorithm departs from the ideas of [50, 72, 60, 59, 58] and proposes computation of all real solutions by numerically solving some non-linear equations. However, it is important to note that the algorithm of [49] is rather specific for rational interpolation problems, while the method of [5, 50, 72] can be naturally extended to the setting of more general moment and analytic interpolation problems, as shown in [58], in which η and a can belong to a more general class of continuous functions on $\mathbb{T}$, instead of simply being polynomials. Therefore, it is of interest to investigate applicability of the homotopy continuation method of Enqvist if η is allowed to have zeros on the unit circle. It has already been argued in Chapter 3 and indicated in Example 3.4.1 therein that such an extension seems to be feasible when the solution f is bounded (has no poles on $\mathbb{T}$). This chapter provides further justification for this observation, by showing that the extension is indeed valid, and goes on to cover the case of unbounded solutions as well. Later on in the chapter, the homotopy continuation method is applied to several examples for practical illustration.

The discussion of this chapter is adapted from the paper [73] (joint work with J. B. Moore).

5.2 Analysis and main results

For $z_0, z_1, \ldots, z_n \in \mathbb{D}$ (note that $z_0 = 0$ by our convention), define

$$\alpha_k(z) = \frac{z + z_k}{z - z_k}$$

whenever z_k has multiplicity 1, and

$$\alpha_k(z) = \frac{z + z_k}{z - z_k} \quad \text{and} \quad \alpha_{k+j}(z) = \frac{2z}{(z - z_k)^{j+1}}$$

for $j = 1, \ldots, m-1$ when z_k has multiplicity m and $z_k = z_{k+1} = \ldots = z_{k+m-1}$. The connection between α_k and Problem 5.1.1 lies in the Herglotz representation [65, 14, 58]. In this representation, any solution of Problem 5.1.1 is expressed as:

$$f(z) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(\theta),$$

where μ is a non-decreasing function of bounded variation on $[-\pi, \pi]$, called the *spectral distribution* of f. The spectral distribution has the decomposition $\mu = \mu_a + \mu_s$, where μ_a is absolutely continuous while μ_s is a piecewise constant function with at most $n - 1$ jumps. This allows us to write each interpolation condition in integral form:

$$\begin{aligned} f(z_k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\theta} + z_k}{e^{i\theta} - z_k} d\mu(\theta) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \alpha_k(e^{i\theta}) d\mu(\theta) = w_k \end{aligned}$$

and

$$\begin{aligned} \frac{1}{j!} f^{(j)}(z_k) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{2e^{i\theta}}{(e^{i\theta} - z_k)^{j+1}} d\mu(\theta) \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} \alpha_{k+j}(e^{i\theta}) d\mu(\theta) \\ &= w_{k+j}, \quad \text{for } j = 0, 1, \ldots, m-1 \end{aligned}$$

whenever $z_k = z_{k+1} = \ldots = z_{k+m-1}$. By a *generalized pseudopolynomial* we mean a complex function of the form $f(z) = a_0 + \sum_{k=1}^{n} (a_k^* \alpha_k + a_k \alpha_{k*})$, where $0 \leq n < \infty$ and $(a_0, a_1, \ldots, a_n) \in \mathbb{R} \times \mathbb{C}^n$. The *order* or *degree* of the generalized

pseudopolynomial f is defined as the largest k such that $a_k \neq 0$ (thus the order is zero if f is a constant function). $\mathfrak{Q}(n, A)$ denotes the set of all generalized pseudopolynomials of order *at most* n with $(a_0, a_1, \ldots, a_n) \in \mathbb{R}\times A^n$, where $A \subseteq \mathbb{C}$. We induce a topology on this set by the $\|\cdot\|_\infty$ norm on the unit circle: $\|f\|_\infty = \operatorname{ess\,sup}_{z\in\mathbb{T}} |f(z)|$, $f \in \mathfrak{Q}(n, A)$. Since α_k has no poles on the unit circle, it can be seen that $\|f\|_\infty$ is well-defined for all $f \in \mathfrak{Q}(n, A)$. We also define $\mathfrak{Q}_+(n, A)$ to be the subset of elements of $\mathfrak{Q}(n, A)$ which are strictly positive (> 0) on $\mathbb{T}$. The restriction of any element of $\overline{\mathfrak{Q}_+(n, A)}\backslash\{0\}$ to $\mathbb{T}$ is a rational spectral density of McMillan degree at most $2n$, thus we shall often view any such element as a spectral density instead of a generalized pseudopolynomial. Hence, to each $d \in \overline{\mathfrak{Q}_+(n, A)}\backslash\{0\}$ we may associate a *unique* outer rational function (i.e., having no roots and poles in $\mathbb{D}$) of McMillan degree at most n, denoted by $\phi(d)$, which is the unique canonical spectral factor of d satisfying: $\phi(d)(0) > 0$ and $|\phi(d)(z)|^2 = d(z)\ \forall z \in \mathbb{T}$.

Let $\tau(z) = \Pi_{k=0}^n(1 - z_k^* z)$ and $H_n = \operatorname{span}\{1, \alpha_{1*}, \ldots, \alpha_{n*}\}$. It will later prove useful to note that H_n has an equivalent description as $H_n = \{f \mid f = \frac{\sigma}{\tau}, \sigma \text{ is a polynomial of degree at most } n\}$ [48]. Then, by definition, any $f \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$ can be written as $f = g + g_*$ with $g \in H_n \cap \mathcal{C}$. Letting $g = \frac{\sigma}{\tau}$, we have that $f = \frac{\sigma\tau_* + \sigma_*\tau}{\tau_*\tau}$ and by spectral factorization of the numerator we may write $f = \frac{\xi_*\xi}{\tau_*\tau}$ for some outer polynomial ξ with $\deg(\xi) \leq n$. Therefore, $\overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\} = \{f \mid f = \frac{\xi_*\xi}{\tau_*\tau},\ \xi \text{ is some outer polynomial, } \deg(\xi) \leq n\}$.

Define the mapping $Q : \mathbb{R} \times \mathbb{C}^n \to \mathfrak{Q}(n, \mathbb{C})$ by:

$$Q(q_0, q_1, q_2, ..., q_n)(z) = q_0 + \sum_{k=1}^{n} \frac{1}{2}(q_k^* \alpha_k + q_k \alpha_{k*}). \tag{5.2}$$

Clearly, Q is a bijective map. Let $\Psi = \frac{\eta_*\eta}{\tau_*\tau}$ with η being a polynomial as defined in Theorem 5.1.2. Then $\Psi \in \overline{\mathfrak{Q}_+(n, \mathbb{C})}\backslash\{0\}$. Let us now consider a modification of the functional $\mathbb{J}_\Psi$ we had encountered in the preceding chapters, which we shall again denote by $\mathbb{J}_\Psi$. This modified functional $\mathbb{J}_\Psi : \overline{Q^{-1}(\mathfrak{Q}_+(n, \mathbb{C}))} \to \mathbb{R} \cup \{\infty\}$ is defined by:

$$\mathbb{J}_\Psi(q) = \Re\big\{w^* q - \langle \Psi, \log Q(q)\rangle\big\}, \tag{5.3}$$

where $w = [w_0\ w_1\ \ldots\ w_n]^T$, $q = [q_0\ q_1\ \ldots\ q_n]^T$ and $\langle f, g\rangle = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(e^{i\theta}) g(e^{i\theta})^* d\theta$.

The (modified) functional was first introduced and its properties studied for $\Psi \in \mathfrak{Q}_+(n, \mathbb{C})$ in [14]. As before, these properties continue to hold for $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$ and they are same ones as given in Theorem 3.3.2 with $\mathbb{J}_\Psi$ now as being defined by (5.3). It is important to note that although Chapters 3 and 4 treat the the rational covariance extension problem, where $z_0 = z_1 = \ldots = z_n = 0$

and $\alpha_k(z) = \frac{2}{z^k}$ for $k = 0, 1, \ldots, n$, it can be seen that since $\overline{\mathfrak{Q}_+(n,\mathbb{C})}$ lies in a finite dimensional space (i.e., $\text{span}\{\alpha_1, \ldots, \alpha_n\} \bigoplus H_n$) and contains functions continuous on $\mathbb{T}$, the analysis therein carries over *mutatis mutandis* to the current setting without technical difficulty (one simply substitutes z^k with $\alpha_{k*}(z)$ and c_k with w_k, etc).

For any $q' \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$, let $\nabla_{q'}\mathbb{J}_\Psi(q)$ again denote the directional derivative of $\mathbb{J}_\Psi$ at the point q in the direction of $q' - q$, i.e.,

$$\nabla_{q'-q}\mathbb{J}_\Psi(q) = \lim_{h \downarrow 0} \frac{\mathbb{J}_\Psi(q + h(q' - q)) - \mathbb{J}_\Psi(q)}{h}. \tag{5.4}$$

Then we again have that $\mathbb{J}_\Psi$ has a unique minimizer q_{min} which is stationary (i.e., $\nabla_{q'-q}\mathbb{J}_\Psi(q_{min}) = 0 \, \forall q' \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$) and lies in $Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$ whenever Ψ is positive definite on $\mathbb{T}$ [6]. It follows that b in Theorem 5.1.2 is given by $b = \tau\phi(Q(q_{min}))$ and a can be found by solving the equation $a_*b + ba_* = \Psi$ [14]. As for the case where Ψ has zeros on $\mathbb{T}$, we have the following generalized version of Theorem 4.2.2:

Theorem 5.2.1 *Let η be as in Theorem 5.1.2, $\Psi = \frac{\eta_*\eta}{\tau_*\tau} \in \partial\mathfrak{Q}_+(n,\mathbb{C})\backslash\{0\}$ and $q_{min} = \underset{q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}}{\arg\min} \mathbb{J}_\Psi(q)$. Then:*

1. *$\nabla_{q'-q}\mathbb{J}_\Psi(q_{min}) = 0$ for all $q' - q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ if and only if the pair (a, b) as defined in Theorem 5.1.2 is such that $f = \frac{a}{b} \in \mathcal{H}^\infty$.*

2. *$q_{min} \in \partial\mathfrak{Q}_+(n,\mathbb{C})$ and $\nabla_{q'-q}\mathbb{J}_\Psi(q_{min}) > 0$ for all $q' \in Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))$ if and only if the pair (a, b) as defined in Theorem 5.1.2 is such that $f = \frac{a}{b}$ has a pole on $\mathbb{T}$. $\nabla_{q'-q}\mathbb{J}_\Psi(q_{min})$ is then given by:*

$$\nabla_{q'-q}\mathbb{J}_\Psi(q_{min}) \;=\; \sum_{l=0}^{m} K_l \Re\Big\{\sum_{k=0}^{n} \alpha_{k*}(e^{i\theta_l})(q'_k - q_{min,k})\Big\}, \tag{5.5}$$

   where $m < n$, $K_0, K_1, \ldots, K_m$ are some positive constants and $\theta_0, \theta_1, \ldots, \theta_m \in (-\pi, \pi]$, with $\theta_i \neq \theta_j$ whenever $i \neq j$, are the discontinuity points of the spectral distribution of f, i.e., $e^{i\theta_0}, \ldots, e^{i\theta_m}$ are the poles of f on $\mathbb{T}$.

Moreover, in both cases $\frac{b}{\tau} = \phi(Q(q_{min}))$ and all roots of $Q(q_{min})$ on $\mathbb{T}$, including multiplicities, are also roots of Ψ.

Proof. Although the proof is analogous to that of Theorem 4.2.2, for the sake of clarity we shall here just detail a possibly not so obvious part in the adaptation of the latter proof needed to establish Point 2 of the theorem. To this end, as in Chapter 4, let us write $f = f_a + f_s$, where $f_a \in \mathcal{C} \cap \mathcal{H}^\infty$ while $f_s \in \mathcal{C}$ has one or more simple poles on $\mathbb{T}$. We also have the representation $f_a(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} d\mu_a(\theta)$ and $f_s(z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} d\mu_s(\theta)$, where μ_a and μ_s are, respectively, the absolutely continuous and singular part of the spectral distribution μ of f. Since $d\mu_a(\theta) = \Re\{f_a(e^{i\theta})\}d\theta$ and $d\mu_s(\theta) = \sum_{l=0}^{m} K_l \delta(\theta - \theta_l) d\theta$ for some positive constants $K_0, K_1, \ldots, K_n$ ($\delta(x)$ denotes the Dirac delta function), we have that $f_a(z_k) = \langle f_a + f_{a*}, \alpha_{k*} \rangle$ and $f_s(z_k) = \sum_{l=0}^{m} K_l \alpha_k(e^{\theta_l})$ (with obvious modification if z_k is a repeated interpolation point). Thus, we obtain the relation $f(z_k) - f_a(z_k) = w_k - \langle f_a + f_{a*}, \alpha_{k*} \rangle = \sum_{l=0}^{m} K_l \alpha_k(e^{i\theta_l})$, in analogy with that obtained in Chapter 4 for the case $z_0 = z_1 = \ldots = z_n = 0$. The relation is a key one for establishing (5.5). The remaining arguments are then straightforward to adapt from the proof of Theorem 4.2.2. □

An important conclusion to be drawn from Theorem 5.2.1 is that, regardless of whether Ψ has zeros on $\mathbb{T}$ or not, the polynomial b of Theorem 5.1.2 associated with Ψ is always given by $b = \tau\phi(Q(q_{min}))$. Once b is computed, a can be obtained by multiplying the coefficients of b by a certain matrix W which only depends on the interpolation data $(z_0, w_0), (z_1, w_1), \ldots, (z_n, w_n)$ (see, e.g., [18, 49] for further details), i.e., if $a(z) = [1\ z\ \ldots\ z^n][a_0\ a_1\ \ldots\ a_n]^T$ and $b(z) = [1\ z\ \ldots\ z^n][b_0\ b_1\ \ldots\ b_n]^T$ then:

$$[a_0\ a_1\ \ldots\ a_n]^T = W[b_0\ b_1\ \ldots\ b_n]^T. \tag{5.6}$$

The only discrepancy is that when Ψ has zeros on $\mathbb{T}$, $\mathbb{J}_\Psi$ may have a minimizer which is *not* a stationary point.

As we have remarked in previous chapters, although properties of $\mathbb{J}_\Psi$ make it convenient for analysis, it is not suitable for numerical optimization, especially when q_{min} is close to or on the boundary. This is due to the fact that the condition number of the Hessian of $\mathbb{J}_\Psi$ tends to ∞ as q_{min} goes to the boundary of $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$. Define

$$D(d)(z) = \sum_{k=0}^{n} d_k z^k$$

and $\mathcal{D}_n = \{d = (d_0, d_1, \ldots, d_n) \in \mathbb{R} \times \mathbb{C}^n \mid d_0 > 0, D(d) \text{ is outer}\}$. Then for $\Psi \in \mathfrak{Q}_+(n,\mathbb{C})$, one way to circumvent the difficulty with $\mathbb{J}_\Psi$, developed in [50, 72, 60, 59, 58], is to reformulate the optimization problem. Recalling from earlier

that every $Q(q) \in \overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$ can be written as $Q(q) = \frac{D(d)_* D(d)}{\tau_* \tau}$ for some $d \in \mathcal{D}_n$, introduce the modified functional $\mathcal{J}_\Psi : \mathcal{D}_n \to \mathbb{R} \cup \{\infty\}$:

$$\mathcal{J}_\Psi(d) = d^* K d - \left\langle \log \frac{D(d)_* D(d)}{\tau_* \tau}, \Psi \right\rangle.$$

where $d = [d_0\, d_1\, \ldots\, d_n]^T$ and K is a positive definite Hermitian matrix which depends only on the interpolation data $\{(z_k, w_k)\}_{k=0,1,\ldots,n}$; an expression for K in terms of $z_0, z_1, \ldots, z_n$ and $w_0, w_1, \ldots, w_n$ are given in [58, 72, 60, 59]. Observe that $\mathcal{J}_\Psi$ can be rewritten as:

$$\mathcal{J}_\Psi(d) = d^* K d - 2\Re\{\langle \log D(d), \Psi \rangle\} + 2\Re\{\langle \log \tau, \Psi \rangle\},$$

where the last term does not depend on d and is not essential in the ensuing analysis. The main idea is, instead of minimizing $\mathbb{J}_\Psi$ over $\overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$, we now minimize $\mathcal{J}_\Psi$ over $\mathcal{D}_n$.

It has been argued in [50], that the new functional is much better suited for numerical treatment as the hessian and its condition number does not blow up as d goes to the boundary of $\mathcal{D}_n$. However, the modified optimization problem is no longer convex since $\mathcal{D}_n$ is not a convex set. Fortunately, due to the bijective correspondence between $\overline{\mathfrak{Q}_+(n,\mathbb{C})}\backslash\{0\}$ and $\mathcal{D}_n$, $\mathcal{J}_\Psi$ has a unique global minimum, and it has been shown that it is locally convex around the global minimum. This makes it possible to find the global minimum of $\mathcal{J}_\Psi$ by constructing a convex homotopy and solving a sequence of locally convex optimization problems as detailed in [50, 72, 60, 59]. We have the following new result which has only been shown previously for $\Psi \in \mathfrak{Q}_+(n,\mathbb{C})$:

Lemma 5.2.2 *For $\Psi \in \partial\mathfrak{Q}_+(n,\mathbb{C})\backslash\{0\}$, $\mathcal{J}_\Psi$ again has a unique minimizer on $\mathcal{D}_n$. Moreover, this minimizer is also stationary.*

Proof. Let s denote the bijective map that sends $a \in \mathcal{D}_n$ to $Q^{-1}(\frac{D(a)_* D(a)}{\tau_* \tau}) \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}\backslash\{0\}$ and note the relation

$$\mathcal{J}_\Psi(a) = \mathbb{J}_\Psi(s(a)).$$

Let q_{min} be as in Theorem 5.2.1 and define $\hat{d} = s^{-1}(q_{min})$. Using the fact that $\mathbb{J}_\Psi(q_{min}) \leq \mathbb{J}_\Psi(q)\ \forall q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}$ (by Theorem 3.3.2), we then have that

$$\begin{aligned}
\mathcal{J}_\Psi(\hat{d}) &= \mathbb{J}_\Psi(q_{min}) \\
&< \mathbb{J}_\Psi(q) \quad \forall q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}\backslash\{q_{min}\} \\
&= \mathcal{J}_\Psi(s^{-1}(q)) \quad \forall q \in \overline{Q^{-1}(\mathfrak{Q}_+(n,\mathbb{C}))}\backslash\{q_{min}\}.
\end{aligned}$$

Therefore, $\mathcal{J}_\Psi(\hat{d}) < \mathcal{J}_\Psi(d)$ for all $d \in \mathcal{D}_n \backslash \{\hat{d}\}$, implying that $\hat{d}$ is the unique minimizer of $\mathcal{J}_\Psi$. This proves the first part of the lemma.

Define the directional derivative of $\mathcal{J}_\Psi$ in the direction of $d' - d$ analogously to (6.12) and denote it by $\nabla_{d'-d}\mathcal{J}_\Psi$, where $d' \in \mathcal{D}_n$. Note that since $\mathcal{D}_n$ is not a convex set, $\nabla_{d'-d}\mathcal{J}_\Psi(d)$ is only defined for *feasible* d' (i.e., defined as those $d' \in \mathcal{D}_n$ for which $(1-h)d + hd' \in \mathcal{D}_n$ for all $0 < h \leq 1$). Let $\mathcal{B}_\Psi = \{d \in \mathcal{D}_n \mid \operatorname{ess\,sup}_{z\in\mathbb{T}} \frac{\Psi(z)}{|D(d)(z)|^2} < \infty\}$. Then, by similar arguments to the proof of Theorem 3.3.7, we may show that $\nabla_{d'-d}\mathcal{J}_\Psi(d)$, $d \in \mathcal{B}_\Psi$, is given by:

$$\nabla_{d'-d}\mathcal{J}_\Psi(d) \;=\; 2\Re\Big\{d^* K(d'-d) - \sum_{i=0}^{n} \big\langle \frac{g_i}{D(d)}, \Psi \big\rangle (d'_i - d_i)\Big\}, \tag{5.7}$$

where $g_i(z) = z^i$. Let d_{min} be the unique minimizer of $\mathcal{J}_\Psi$. Since $d_{min} = s^{-1}(q_{min})$, it follows that $D(d_{min}) = \tau\phi(Q(q_{min}))$ and $d_{min} \in \mathcal{B}_\Psi$. Now, let Ψ_k, $k = 1, 2, \ldots$, be a sequence such that $\Psi_k \in \mathfrak{Q}_+(n, \mathbb{C})$ for all k and Ψ_k converges to Ψ uniformly on $\mathbb{T}$, i.e., $\lim_{k\to\infty} \|\Psi - \Psi_k\|_\infty = 0$ and let $d^k_{min} = \arg\min_{d\in\mathcal{D}_n} \mathcal{J}_{\Psi_k}(d)$. Then as shown in the proof of Theorem 4.2.2 :

$$\lim_{k\to\infty} \|d_{min} - d^k_{min}\|_2 = 0 = \lim_{k\to\infty} \|D(d_{min}) - D(d^k_{min})\|_\infty.$$

Furthermore, it has been shown (see, e.g., [58]) that $\nabla_{d'-d}\mathcal{J}_{\Psi_k}(d^k_{min}) = 0$ for all feasible $d' \in \mathcal{D}_n$ and for all k. Now, by the uniform convergence of Ψ_k to Ψ and $D(d^k_{min})$ to $D(d_{min})$ as noted above, we have:

$$\lim_{k\to\infty} \frac{\Psi_k(z)}{D(d^k_{min})(z)} = \frac{\Psi(z)}{D(d_{min})(z)} \quad \text{for a.a. } z \in \mathbb{T}, \tag{5.8}$$

with the exceptional points being the roots of $D(d_{min})$ on $\mathbb{T}$ (which are also roots of Ψ by Theorem 5.2.1). Since $\Psi_k = D(d^k_{min})D(Wd^k_{min})_* + D(d^k_{min})_* D(Wd^k_{min})$ (see Eq. (5.6) and the discussion on the associated page), we have that

$$\begin{aligned}
\Big\|\frac{\Psi_k}{D(d^k_{min})}\Big\|_\infty &= \Big\|D(Wd^k_{min})_* + \frac{D(d^k_{min})_*}{D(d^k_{min})} D(Wd^k_{min})\Big\|_\infty \\
&\leq \|D(Wd^k_{min})_*\|_\infty + \Big\|\frac{D(d^k_{min})_*}{D(d^k_{min})}\Big\|_\infty \|D(Wd^k_{min})\|_\infty \\
&= 2\|D(Wd^k_{min})\|_\infty.
\end{aligned}$$

Now, since $D(Wd^k_{min}) \overset{\|\cdot\|_\infty}{\to} D(Wd_{min})$ as $k \to \infty$, it follows that

$$\sup_{k\geq 1} \|D(Wd^k_{min})\|_\infty < \infty.$$

Consequently,

$$\sup_{k\geq 1}\Big\|\frac{\Psi_k}{D(d^k_{min})}\Big\|_\infty < \infty, \tag{5.9}$$

i.e., the sequence $\{\|\frac{\Psi_k}{D(d^k_{min})}\|_\infty; k = 1, 2, \ldots\}$ is uniformly bounded. Now, by plugging (5.8) into (5.7), and invoking the Lebesque Dominated Convergence Theorem [54] by using (5.9), we get:

$$\begin{aligned}
\nabla_{d'-d}\mathcal{J}_\Psi(d_{min}) &= \lim_{k\to\infty} 2\Re\Big\{(d^k_{min})^* K(d' - d^k_{min}) - \\
&\qquad \sum_{i=0}^{n}\big\langle\frac{g_i}{D(d^k_{min,i})}, \Psi_k\big\rangle(d'_i - d^k_{min,i})\Big\} \\
&= \lim_{k\to\infty} \nabla_{d'-d}\mathcal{J}_{\Psi_k}(d^k_{min}) \\
&= \lim_{k\to\infty} 0 = 0 \quad \text{for all feasible } d' \in \mathcal{D}_n.
\end{aligned}$$

This shows that d_{min} is a stationary point and completes the proof of the lemma. □

Lemma 5.2.2 shows a *striking difference* between $\mathcal{J}_\Psi$ and $\mathbb{J}_\Psi$: for $\Psi \in \partial\mathfrak{Q}_+(n, \mathbb{C})\backslash\{0\}$, the minimizer of $\mathcal{J}_\Psi$ is *always stationary* while the minimizer of $\mathbb{J}_\Psi$ may not be. From the lemma the following is easily obtained:

Corollary 5.2.3 *The functional $\mathcal{J}_\Psi$ is locally strictly convex in a neighborhood of its unique minimizer.*

Proof. Again, let d_{min} denote the unique minimizer of $\mathcal{J}_\Psi$ and let $\mathcal{B}_\Psi$ be defined as before. Recall that $d_{\min} \in \mathcal{B}_\Psi$ since all roots of $D(d_{min})_* D(d_{min})$ on $\mathbb{T}$, counting multiplicities, are also roots of Ψ. For $d \in \mathcal{B}_\Psi$ and a feasible d', define the function $f_{d,d'}$ on $[0, \|D(d' - d)\|_\infty) \subset \mathbb{R}$ by $f_{d,d'}(x) = \mathcal{J}_\Psi(d + x\frac{d'-d}{\|D(d'-d)\|_\infty})$. Let $\frac{d^2_+}{d_+x^2} f_{d,d'}$ denote the *right sided* second derivative of $f_{d,d'}$. Then we may, analogously as before, show that $\frac{d^2_+}{d_+x^2} f_{d,d'}$ is given by (recall that $g_i(z) = z^i$):

$$\begin{aligned}
\frac{d^2_+}{d_+x^2} f_{d,d'}(x) &= 2\Re\Big\{\frac{(d' - d)^* K(d' - d)}{\|D(d' - d)\|^2_\infty} + \\
&\qquad \sum_{i=0}^{n}\sum_{j=0}^{n}\big\langle\frac{g_i g_j}{D(d + x\frac{d'-d}{\|D(d'-d)\|_\infty})^2}, \Psi\big\rangle \times \\
&\qquad \frac{(d'_i - d_i)(d'_j - d_j)}{\|D(d' - d)\|^2_\infty}\Big\}.
\end{aligned}$$

In particular, $\frac{d_+^2}{d_+x^2} f_{d,d'}(0)$ exists and is bounded for all feasible $d' \in \mathcal{D}_n$. Since d_{min} is the unique stationary minimizer of $\mathcal{J}_\Psi$ (by the lemma) and $\frac{d_+^2}{d_+x^2} f_{d_{min},d'}$ is continuous on $[0, \delta)$ for some $\delta > 0$, we must have that $\frac{d_+^2}{d_+x^2} f_{d_{min},d'}$ is positive definite on $[0, \delta')$ for some $0 < \delta' \leq \delta$. Hence, $\mathcal{J}_\Psi$ is strictly convex on any sufficiently small convex subset of $\mathcal{D}_n$ containing $d_{\min}$. □

Lemma 5.2.2 and Corollary 5.2.3 justify the use of the homotopy continuation method for finding solutions of Problem 1 corresponding to η with spectral zeros on the unit circle. Although the functional is not globally convex, we do have a unique stationary minimizer and local strict convexity around that minimizer. This is enough to allow us to use a homotopy continuation to circumvent the lack of global convexity, and solve a sequence of locally convex problems, as is done for the case where all spectral zeros are strictly inside the unit circle. In the next section, we put our assertions to the test by applying the continuation method to compute the different kinds of possible solutions as summarized in Theorem 5.2.1.

5.3 Numerical examples

In this section we present numerical results from application of the continuation method for computing solutions of Problem 1 with spectral zeros on $\mathbb{T}$. Although our results have been developed for a general case, in the examples we restrict our attention to the rational covariance extension problem, i.e. $z_0 = z_1 = \ldots = z_n = 0$, which is the special problem of interest in this thesis. Moreover, to avoid complex arithmetics, we shall only consider the real case, where $w_0, w_1, \ldots, w_n \in \mathbb{R}$ and $\Psi \in \overline{\mathfrak{Q}_+(n, \mathbb{R})} \backslash \{0\}$. We implement the homotopy continuation algorithm as described in [50] and use the stopping criteria: $e_m = \|\nabla \mathcal{J}_\Psi(\hat{d}_m) - \nabla \mathcal{J}_\Psi(\hat{d}_{m-1})\|_2 < \epsilon$ for a specified tolerance $\epsilon > 0$, where $\nabla \mathcal{J}_\Psi$ denotes the gradient of $\mathcal{J}_\Psi$ and $\hat{d}_m$ denotes the iterate (approximation of d_{min}) at the m-th iteration of the algorithm. In all examples, we take the step size $\rho = 0.1$ (see [50, p. 1196]) and set $\epsilon = 10^{-8}$. The computations were executed in Matlab with double precision, but to avoid unnecessary clutter, we shall only display the numerical results up to four digits behind the decimal. In the following, η has been chosen in accordance with Theorem 5.1.2.

Example 5.3.1 *Let the true $f \in \mathcal{C}$ be:*

$$f(z) = \frac{(z-2)(z-1)}{(z-4)(z-5)}.$$

Then $w_0 = 0.2$, $w_1 = -0.1050$ *and* $w_2 = -0.0023$. *Suppose we choose* $\eta(z) = (z-1)(z-0.74053618)$ *having a root at* $z = 1$. *The algorithm returns* $d_{min} = (3.6694, -1.6512, 0.1835)$. *Whence,* $b(z) = 3.6694 - 1.6512z + 0.1835z^2$ *and* $a(z)$ *can be computed to be* $a(z) = 0.3669 - 0.5504z + 0.1835z^2$. *The computed solution* $\hat{f}$ *is:*

$$\begin{aligned}\hat{f}(z) &= \frac{0.3669 - 0.5504z + 0.1835z^2}{3.6694 - 1.6512z + 0.1835z^2} \\ &= \frac{2 - 3z + z^2}{20 - 9z + z^2},\end{aligned}$$

which happens to coincide with the true solution. The example illustrates the case where there are spectral zeros on the unit circle, yet b *is in the interior of* $\mathcal{D}_n$ *(cf. Point 1 of Theorem 5.2.1). Another example of this type had also been given in Example 3.3.4 of Chapter 3 but was computed using the convex optimization approach of [5].*

Example 5.3.2 *Let the true* $f \in \mathcal{C}$ *be:*

$$f(z) = \frac{z-2}{z-4}.$$

Then $w_0 = 1$, $w_1 = -0.125$, *and* $w_2 = -0.0313$. *Now, we choose* $\eta(z) = (z+1)(z-0.38196601)$. *The algorithm returns* $d_{min} = (1.0093, 0.7569, -0.2523)$. *Thus,* $b(z) = 1.0093 + 0.7569z - 0.2523z^2$ *and* $a(z)$ *can be computed to be* $a(z) = 0.5046 + 0.2523z - 0.2523z^2$. *Thus, the computed solution is:*

$$\begin{aligned}\hat{f}(z) &= \frac{0.5046 + 0.2523z - 0.2523z^2}{1.0093 + 0.7569z - 0.2523z^2} \\ &= \frac{-2 - z + z^2}{-4 - 3z + z^2} = \frac{z-2}{z-4}.\end{aligned}$$

It may be inspected that both a *and* b *has one root near* -1 *(which should cancel in the absence of numerical errors). This example serves to illustrate the case where* a, b *and* Ψ *all share a root on* $\mathbb{T}$ *and* $\frac{a}{b} \in \mathcal{H}^\infty$. *Note that another example of this type had been given earlier in Example 3.3.5 of Chapter 3.*

Example 5.3.3 *Let the true* $f \in \mathcal{C}$ *be:*

$$f(z) = \frac{z-2}{z-4} + \frac{1}{4}\frac{1+e^{i\frac{\pi}{6}}z}{1-e^{i\frac{\pi}{6}}z} + \frac{1}{4}\frac{1+e^{-i\frac{\pi}{6}}z}{1-e^{-i\frac{\pi}{6}}z},$$

and note how f *has two simple poles on* $\mathbb{T}$. *Then* $w_0 = 2$, $w_1 = 0.7410$, $w_2 = 0.4687$, *and* $w_3 = -0.0078$. *Set* $\eta(z) = (z - e^{i\frac{\pi}{6}})(z - e^{-i\frac{\pi}{6}})(z - 0.38196601)$.

The algorithm returns $d_{min} = (1.0092, -2.0004, 1.4463, -0.2523)$. *Thus,* $b(z) = 1.0092 - 2.0004z + 1.4463z^2 - 0.2523z^3$, $a(z) = 0.5046 - 1.1263z + 0.9416z^2 - 0.2523z^3$ *and the computed solution* $\hat{f}$ *is:*

$$\hat{f}(z) = \frac{0.5046 - 1.1263z + 0.9416z^2 - 0.2523z^3}{1.0092 - 2.0004z + 1.4463z^2 - 0.2523z^3}.$$

It may be inspected that b *has roots almost at* 4, $e^{i\frac{\pi}{6}}$ *and* $e^{-i\frac{\pi}{6}}$, *but* a *does not. The example serves to illustrate the case where the true solution* f *is an unbounded solution with poles on* $\mathbb{T}$ *(cf. Point 2 of Theorem 5.2.1).*

5.4 Concluding remarks

In this chapter we have shown that a certain homotopy continuation method, originally due to Enqvist, for computing solutions of degree constrained rational interpolation problems with strictly positive parametrizing functions, remains applicable when the parametrizing function is non-strictly positive definite (i.e., have zeros on the unit circle). This includes, as a special case, solutions with one or more poles on the unit circle. It was not previously known that this method can handle such cases, although it has been observed [72, 59] that it is able to compute solutions with poles very close to the unit circle with high accuracy.

A potential advantage of the homotopy continuation method over the alternative method proposed in [49] is that of *generality*. Indeed, it has already been adapted for computing strictly positive and absolutely continuous solutions of more general moment and analytic interpolation problems with a complexity constraint [58]. Therefore, generalizations of the developments in this chapter may allow for computation of non-strictly positive and non-absolutely continuous solutions of these more general problems. Moreover, in [74, 75] fast algorithms have recently been proposed for the original convex optimization approach of [5, 14] and, because of some "structural similarities" between that approach and the continuation method, it would be interesting to investigate if similar fast algorithms can be developed to solve each of the local convex optimization problems in the continuation method. These are some topics which can be considered for future research.

Chapter 6

Spectral Factorization of a Class of Matrix-Valued Spectral Densities

6.1 Introduction

In this chapter we now develop a new approach to spectral factorization as motivated and outlined in Chapter 1, Section 1.1, using the ideas and convex optimization techniques from Chapters 3 and 4.

Instead of directly approximating a spectral factor as with the methods described in [64], the strategy employed here is to construct a rational approximation of the spectral density and perform spectral factorization on the approximate spectral density to obtain a rational shaping filter. However, the latter spectral factorization need not be performed separately, but becomes part of the proposed procedure thanks to the continuation method described in Chapter 5. The main question here is whether the approximate canonical spectral factor (i.e., the unique spectral factor which is positive at the origin) that is obtained in this way will be a good approximation of the true canonical spectral factor. This question is equivalent to asking whether the operation of taking canonical spectral factors is *continuous*. It has recently been shown that such an operation is *sequentially continuous*: Given a sequence of spectral densities which converge to a limiting spectral density (in the space of functions integrable on the unit circle) then their canonical spectral factors will also converge to that of the limiting spectral density if a uniform log-integrability assumption on the spectral densities is satisfied

[15] (for a related result, see also [76]). This property is then exploited to ensure that the resulting approximate canonical spectral factor is close to the true one in an appropriate norm. In particular, we first derive some easily verified sufficient conditions which guarantee uniform log-integrability of a sequence of spectral densities.

The approximating rational spectral densities are constructed using the theory of degree constrained rational covariance extensions studied in Chapters 3 and 4. Under some mild regularity conditions on a given spectral density, theoretical results will be derived to show that certain covariance matching rational spectral densities, and also their canonical factors, will converge to, respectively, the given spectral density and its canonical spectral factor (in the appropriate vector spaces). Based on this construction a new algorithm is proposed which give freedom in selection of spectral zeros for the approximating spectral densities. Conditions on the selected spectral zeros for convergence of the algorithm in $\mathcal{H}^2$ and $\mathcal{H}^\infty$ will be given as well as a heuristic scheme for their selection.

In the penultimate section of this chapter, several simulations are executed in order to compare the performance of the new algorithm over the popular maximum entropy method for spectral factorization of possibly non-rational spectral densities having one or more zeros on or close to the unit circle. As discussed in Section 1.1, the maximum entropy method, while being able to handle quite a general class of spectral densities, suffers from slow convergence when the spectral density is non-coercive [12, 9, 10]. This can lead to approximate rational canonical spectral factors of unnecessarily high degree because each iteration increases the degree of the approximation. The comparative simulations indicate advantages of the new algorithm offers over the maximum entropy approach: lower degree approximations with lower approximation error (defined in a certain sense). In particular, in two simulations we successfully construct approximate rational canonical spectral factors for the non-rational and non-coercive Kolmogorov and von Karman spectral densities which are of interest in the study of atmospheric and wind turbulence.

This chapter is organized as follows. In Section 6.2 we introduce some additional notation and recall some definitions and results from the literature. Following that, in Section 6.3 we discuss a result on sequential continuity of the spectral factorization mapping. In Section 6.4 we derive a new set of easily checkable and sufficient conditions for uniform log-integrability of a sequence of spectral densities. In Sections 6.5 and 6.6 we develop the theoretical foundation of a new

approach to spectral factorization and introduce a new spectral factorization algorithm for a class of matrix-valued spectral densities. We then present a number of numerical examples using the proposed spectral factorization algorithm in Section 6.7. Finally in Section 6.8 we give the conclusions of this chapter and discuss potential applications of the results as well as directions for future research.

The discussion of this chapter is adapted from the papers [77, 66].

6.2 Additional notation and definitions

First we introduce some additional notation which will be required for this chapter and generalize some of the definitions from previous chapters as well as recalling some definitions and relevant results from the literature.

- $\Re\{A\} = A + A^*$ denote the hermitian transpose and hermitian part of a complex matrix A, respectively.
- A *pseudopolynomial* is a $\mathbb{C}^{l\times l}$-valued (with $l \in \mathbb{N}$) function f of the form $f(z)= \sum_{i=-m}^{n} A_i z^i$, where $0 \leq m, n < \infty$ and $A_i \in \mathbb{C}^{l\times l}$ for $i = -m, \ldots, n$.
- The $\|\cdot\|_p$ norm of a matrix $A \in \mathbb{C}^{m\times n}$ is defined as [15]:

$$\|A\|_p = \begin{cases} \left(\mathrm{Tr}\{(A^*A)^{p/2}\}\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty, \\ \sup_{v\in\mathbb{C}^n, \|v\|\leq 1} \|Av\| & \text{if } p = \infty. \end{cases}$$

- μ denotes the Lebesque measure on $\mathbb{T}$.
- $\mathcal{L}^p_{m\times n}$, $1 \leq p \leq \infty$, denotes the space of measurable functions mapping from $\mathbb{T}$ to $\mathbb{C}^{m\times n}$ with a finite $\|\cdot\|_p$ norm defined by:

$$\|f\|_p = \begin{cases} \left(\frac{1}{2\pi}\int_{\mathbb{T}} \|f(z)\|_p^p d\mu\right)^{\frac{1}{p}} & \text{if } 1 \leq p < \infty \\ \operatorname{ess\,sup}_{z\in\mathbb{T}} \|f(z)\|_\infty & \text{if } p = \infty \end{cases}$$

  If $n = 1$, we write $\mathcal{L}^p_{m\times n}$ simply as $\mathcal{L}^p_m$.
- $\mathcal{H}^p_{m\times n}$, $1 \leq p \leq \infty$, denotes the subspace of functions in $\mathcal{L}^p_{m\times n}$ having an analytic continuation from $\mathbb{T}$ to $\mathbb{D}$. If $n = 1$, we write $\mathcal{H}^p_{m\times n}$ simply as $\mathcal{H}^p_m$.
- H_* denotes the *parahermitian conjugate* of a $\mathbb{C}^{m\times n}$-valued complex function H: $H_*(z) = H(z^{*-1})^*$.

If H is a rational element of $\mathcal{H}^p_{n\times n}$ or $\mathcal{L}^p_{n\times n}$ then the degree of H, denoted by $\deg(H)$, is defined to be the McMillan degree of H. Let $\mathcal{P}_n$ denote the linear space of $\mathbb{C}^n-$valued trigonometric polynomials on $\mathbb{T}$. It is well-known that this space is dense in $\mathcal{L}^p_n$ for all $p \in [1,\infty)$. In a similar fashion we define the linear space $\mathcal{P}^+_n$ to be the set of $\mathbb{C}^n-$valued polynomials on $\mathbb{C}$. We may view $\mathcal{P}^+_n$ as a linear subspace of $\mathcal{P}_n$. A function $\rho \in \mathcal{H}^2_{n\times n}$ is said to be *outer* if $\overline{\rho\mathcal{P}^+_n} = \mathcal{H}^2_n$, i.e., the set of products $\rho\mathcal{P}^+_n$ is dense in $\mathcal{H}^2_n$ [15]. In the special case where $n = 1$ (the scalar case) and ρ is a rational function, it is known that ρ is outer if and only if all its zeros and poles lie in $\mathbb{C}\backslash\mathbb{D}$.

A function W which maps from $\mathbb{T}$ to $\mathbb{C}^{n\times n}$ is a *spectral density* if 1) it is in $\mathcal{L}^1_{n\times n}$, and 2) there exists an outer function $H \in \mathcal{H}^2_{n\times n}$ such that $W(e^{i\theta}) = H(e^{i\theta})^* H(e^{i\theta})$. Note that the definition implies that $W^* = W$ and W is non-negative definite a.e. $\mathbb{T}$. The function H is called a *spectral factor* of W. A spectral factor can be uniquely specified if a condition is imposed on its value at the origin. We call the *unique* spectral factor which is positive definite at the origin the *canonical spectral factor* (CSF). We say that W is *rational* if each element W_{ij} is of the form $W_{ij}(e^{i\theta}) = \frac{P_{ij}(e^{i\theta})}{Q_{ij}(e^{i\theta})}$ for some scalar pseudopolynomials P_{ij} and Q_{ij}. A precise characterization of spectral densities is given in the following classical result:

Theorem 6.2.1 ([39, 42, 1, 64]) *A non-negative definite function $W \in \mathcal{L}^1_{n\times n}$ is a spectral density if and only if $\int_{\mathbb{T}} |\log\det W(z)|\,\mu(dz) < \infty$.*

For a function $f : \mathbb{T} \to \mathbb{R}$ we write $f > 0$ ($f \geq 0$) if f is positive (non-negative) definite a.e. $\mathbb{T}$, and $f > g$ ($f \geq g$) will be taken to mean $f - g > 0$ ($f - g \geq 0$). A spectral density W is said to be *coercive* if $W \geq \delta > 0$, otherwise it is *non-coercive*.

6.3 Sequential continuity of the spectral factorization mapping

Let W be a spectral density and let $\Phi(W)$ denote its unique CSF. Then the mapping $\Phi : W \mapsto \Phi(W)$ is called the *spectral factorization mapping*. It was recently shown in [15] that the mapping Φ is *sequentially continuous*, that is

Theorem 6.3.1 *Let W be a spectral density, and let $\{W_r\}_{r\in\mathbb{N}}$ be a sequence of spectral densities such that $W_r \to W$ in $\mathcal{L}^1_{n\times n}$ as $r \to \infty$. Then the following are equivalent:*

1. *The sequence $\{\log\det W_r\}_{r\in\mathbb{N}}$ is uniformly integrable.*
2. *$\Phi(W_r) \to \Phi(W)$ in $\mathcal{H}^2_{n\times n}$ as $r \to \infty$.*

Recall that a family of scalar measurable functions $\{X_\gamma \mid \gamma \in \Gamma\}$ parametrized by a non-empty set Γ on a measurable space $(\Omega, \mathcal{F})$ with measure M is said to be *uniformly integrable* if $\lim_{\alpha\to\infty} \sup_{\gamma\in\Gamma} \int_{\{\omega\in\Omega||X_\gamma(\omega)|>\alpha\}} |X_\gamma(\omega)| M(d\omega) = 0$.

Remark 6.3.2 *We shall refer to the condition in Point 1 of Theorem 6.3.1 as uniform log-integrability.*

Several conditions which are equivalent to uniform log-integrability are given in [15, Proposition 4.2]. However, these conditions are general and do not indicate how to construct a uniformly log-integrable sequence $\{W_r\}_{r\in\mathbb{N}}$ which converges to W in $\mathcal{L}^1_{n\times n}$. For this reason, we shall shortly develop more explicit sufficient conditions.

6.4 A sufficient and verifiable set of conditions for uniform log-integrability

In this section we shall derive a new set of conditions on the sequence of convergent spectral densities *and* the limiting spectral density which ensures that the uniform log-integrability condition of Theorem 6.3.1 is satisfied. To this end, for $\alpha \geq 0$, let us define:

$$\begin{aligned} A_r(\alpha) &= \{z \in \mathbb{T} \mid |\log\det W_r(z)| > \alpha\}, \\ A_{r+}(\alpha) &= \{z \in \mathbb{T} \mid \det W_r(z) > e^{\alpha}\}, \\ A_{r-}(\alpha) &= \left\{z \in \mathbb{T} \mid \det W_r(z) < e^{-\alpha}\right\} \end{aligned}$$

and note that $A_{r+}(\alpha) \cap A_{r-}(\alpha) = \phi$ and $A_r(\alpha) = A_{r+}(\alpha) \cup A_{r-}(\alpha)$. The set $A_{r+}(\alpha)$ is the collection of points at which $\det W_r$ has "large" values and which may grow to ∞ as $\alpha \to \infty$, while $A_{r-}(\alpha)$ is the set of points where $\det W_r$ take

on "small" values and can diminish to 0 as $\alpha \to \infty$. Then we have the following inequality:

$$\sup_{r\in\mathbb{N}} \int_{A_r(\alpha)} |\log\det W_r(z)| d\mu \quad \leq \quad \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \log\det W_r(z) d\mu + \sup_{r\in\mathbb{N}} \int_{A_{r-}(\alpha)} -\log\det W_r(z) d\mu. \quad (6.1)$$

The main idea here is to derive sufficient conditions for each of the two terms on the right hand side of (6.1) such that they go to 0 as $\alpha \to \infty$. It turns out that finding conditions to guarantee the desired effect on the first term is relatively easy. As for the second term, the conditions are more complicated. To have that term go to 0 as $\alpha \to \infty$, the idea is to impose conditions which exclude the existence of a set of positive Lebesque measure on which $\det W_r$ decays to zero as $r \to \infty$. Before going into the formal details, we note the following matrix inequality:

Lemma 6.4.1 *For any non-negative definite matrix* $A \in \mathbb{C}^{n\times n}$, $\log\det A \leq \|A\|_1$.

Proof. Note that the result is trivial if A is singular, since in this case we have $\log\det A = -\infty$. Therefore we assume that A is *positive definite*. Let $\sigma_1, \sigma_2, \ldots, \sigma_n$ be the singular values of A, with $\sigma_1 \geq \sigma_2 \geq \ldots > 0$. Since A is positive definite, we have that $\det(A^*) = \det(A)$ and $\log\det A = \frac{1}{2}\log\det(AA^*) = \sum_{k=1}^n \log\sigma_k$. On the other hand, we also have that $\|A\|_1 = \mathrm{Tr}((AA^*)^{\frac{1}{2}}) = \sum_{k=1}^n \sigma_k$ and the result follows since $\log\sigma_k \leq \sigma_k$ for $k = 1, \ldots, n$. □

First, let us make the following assumptions:

A1. $\operatorname{ess\,sup}_{z\in\mathbb{T}} \|W_r(z)\|_1 < \infty$ for all $r \in \mathbb{N}$.

A2. The sequence $\{W_r\}_{r\in\mathbb{N}}$ converges in $\mathcal{L}^1_{n\times n}$ to W as $r \to \infty$.

Now we can show the following result:

Lemma 6.4.2 *Under* Assumption A1-A2*:*

$$\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \log\det W_r(z)\mu(dz) = 0.$$

Proof. By Assumption A1 and A2 we have i) $\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \|W(z) - W_r(z)\|_1 \mu(dz) = 0$ and ii) $\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \mu(A_{r+}(\alpha)) = 0$ (if the latter is not

true then we would have $\sup_{r\in\mathbb{N}} \|W_r\|_1 = \infty$ which contradicts A1 and A2). It then follows from ii that iii) $\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \|W(z)\|_1 \mu(dz) = 0$. Since $\|W_r(z)\|_1 \leq \|W(z) - W_r(z)\|_1 + \|W(z)\|_1$, we get from i and iii that

$$\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \|W_r(z)\|_1 \mu(dz) = 0.$$

Lemma 4.1 then gives $\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r+}(\alpha)} \log\det W_r(z)\mu(dz) = 0$, as desired. □

Let us impose three further assumptions on $\{W_r\}_{r\in\mathbb{N}}$:

A3. $W_r(e^{i\theta})$ is a piecewise continuous function of θ for each $r \in \mathbb{N}$.

A4. Let Z_a be the set defined by:

$$Z_a = \left\{ z_0 \in \mathbb{T} \mid \liminf_r f_r = 0 \ \forall \text{neighborhoods } U \text{ of } z_0, f_r = \inf_{z\in U\cap\mathbb{T}} \det W_r(z) \right\}.$$

Then the cardinality of Z_a is finite.

A5. Let Z_r be the set of all zeros of $\det W_r$ (i.e., all points $z_0 \in \mathbb{T}$ for which $\inf_{z\in U\cap\mathbb{T}} \det W_r(z) = 0 \ \forall \text{neighborhoods } U \text{ of } z_0$). Then $\exists M_1, M_2, \Delta_1, \Delta_2 > 0$ such that for any $r \in \mathbb{N}$ and any $\theta_{0,r} \in (-\pi, \pi]$ such that $e^{i\theta_{0,r}} \in Z_r \cup Z_a$:

$$\det W_r(e^{i\theta}) \geq M_1 \left|\theta - \theta_{0,r}\right|^{M_2} \forall \theta \in [\theta_{0,r} - \Delta_1, \theta_{0,r} + \Delta_2] \cap (-\pi, \pi]. \quad (6.2)$$

Remark 6.4.3 *Assumption A5 implies that the cardinality of Z_r is uniformly bounded (away from ∞).*

We have the following result:

Lemma 6.4.4 *Under* Assumption A3-A5*:*

$$\lim_{\alpha\to\infty} \sup_{r\in\mathbb{N}} \int_{A_{r-}(\alpha)} -\log\det W_r(z)\mu(dz) = 0.$$

Proof. Let $\theta_{r,1}, \ldots, \theta_{r,n_r}$ be the angles (in $(-\pi, \pi]$) of elements of $Z_r \cup Z_a$. Then $n_r \leq L$ for all r, where L is some positive integer. Define:

$$\begin{aligned} \tilde{A}_{r-}(\alpha) &= \left\{ \theta \in (-\pi, \pi] \mid e^{i\theta} \in A_{r-}(\alpha) \right\} \text{ and} \\ \tilde{A}_{r-,k}(\alpha) &= \tilde{A}_{r-}(\alpha) \bigcap \left\{ \theta \in (-\pi, \pi] \mid -\Delta_1 \leq \theta - \theta_{r,k} \leq \Delta_2 \right\}, \end{aligned}$$

for $i = 1, \ldots, n_r$. Note that $\theta_{r,k} \in \tilde{A}_{r-}(\alpha)$ for $k = 1, \ldots, n_r$ and that $\tilde{A}_{r-,k}$ can be empty for some k's. Clearly, Assumption A3-A5 imply that for some α_0 large

enough and $\alpha > \max\left\{\alpha_0, -\log\left(M_1\left(\min\{\Delta_1,\Delta_2\}\right)^{M_2}\right)\right\}$, $\tilde{A}_{r-,k}(\alpha)$ are disjoint for $k = 1,\ldots,n_r$, independently of r, and $\tilde{A}_{r-}(\alpha) = \bigcup_{k=1}^{n_r}\tilde{A}_{r-,k}(\alpha)$. Furthermore, without loss of generality we may take M_1 to have value less than 1. Hence the following holds:

$$\begin{aligned}\int_{A_{r-}(\alpha)} -\log\det W_r(z)d\mu &= \sum_{k=1}^{n_r}\int_{\tilde{A}_{r-,k}(\alpha)} -\log\det W_r(e^{i\theta})d\theta,\\ &\le \sum_{k=1}^{n_r}\int_{\tilde{A}_{r-,k}(\alpha)} -\log\left(M_1|\theta-\theta_{r,k}|^{M_2}\right)d\theta,\\ &\le -\mu(A_{r-}(\alpha))\log M_1\\ &\qquad + M_2\sum_{k=1}^{n_r}\int_{\tilde{A}_{r-,k}(\alpha)} -\log|\theta-\theta_{r,k}|\,d\theta. \quad (6.3)\end{aligned}$$

Let $\alpha_1 = \max\left\{\alpha_0, -\log\left(M_1\left(\min\{\Delta_1,\Delta_2\}\right)^{M_2}\right)\right\}$. Assumption A3-A5 also imply that for $\alpha > \alpha_1$ there exists a number $\epsilon(\alpha) > 0$, dependent on α, such that $\lim_{\alpha\to\infty}\epsilon(\alpha) = 0$ and $\tilde{A}_{r-,k}(\alpha) \subset \theta_{r,k}+B(\alpha) = \{\theta\in(-\pi,\pi] \mid \theta = \theta_{r,k}+\omega;\omega\in B(\alpha)\}$, where $B(\alpha)$ is a set independent of r defined by $B(\alpha) = \{\theta\in(-\pi,\pi] \mid -\epsilon(\alpha)\Delta_1 \le \theta \le \epsilon(\alpha)\Delta_2\}$. Therefore from (6.3) we have:

$$\begin{aligned}\int_{A_{r-}(\alpha)} -\log\det W_r(z)d\mu &\le -\sum_{k=1}^{n_r}\Lambda(\theta_{r,k}+B(\alpha))\log M_1\\ &\qquad + M_2\sum_{k=1}^{n_r}\int_{\theta_{r,k}+B(\alpha)} -\log|\theta-\theta_{r,k}|d\theta,\\ &\le -L\Lambda(B(\alpha))\log M_1 + LM_2\int_{B(\alpha)} -\log|\theta|d\theta,\end{aligned}$$

where Λ denotes the Lebesque measure on $(-\pi,\pi]$. Since the right hand side of the last inequality $\to 0$ as $\alpha\to\infty$ independently of r, we conclude that $\lim_{\alpha\to\infty}\sup_{r\in\mathbb{N}}\int_{A_{r-}(\alpha)} -\log\det W_r(z)\mu(dz) = 0$, which is the statement we had set out to prove. □

A direct consequence of Lemma 6.4.2 and Lemma 6.4.4 is the following theorem:

Theorem 6.4.5 *Under* Assumption A1-A5*:*

$$\lim_{\alpha\to\infty}\sup_{r\in\mathbb{N}}\int_{A_r(\alpha)} |\log\det W_r(z)|\mu(dz) = 0,$$

i.e., the sequence $\{\log\det W_r\}_{r\in\mathbb{N}}$ *is uniformly integrable.*

Proof. Follows directly from Lemma 6.4.2 and Lemma 6.4.4 by taking the limit $\alpha \to \infty$ on both sides of inequality (6.1). □

The above theorem has the following important corollary:

Corollary 6.4.6 *Let $W \in \mathcal{L}^1_{n\times n}$ with $\|\det W - f\|_1 = 0$ for some spectral density $f \in \mathcal{L}^1$ having a finite number of zeros on $\mathbb{T}$ (a zero is as defined in Assumption A5) . If $\{W_r\}_{r\geq 1}$ is a sequence of piecewise continuous spectral densities such that $\lim_{r\to\infty} \operatorname{ess\,sup}_{z\in\mathbb{T}} \|W(z) - W_r(z)\|_1 = 0$ then $\lim_{r\to\infty} \|\Phi(W) - \Phi(W_r)\|_2 = 0$.*

The corollary is a simple but useful result and relaxes the requirement $W > 0$ in [19, Theorem 1] (or [15, Corollary 6.2] with $p = 2$). We shall prove later on, that a sequence satisfying the conditions of the corollary can be explicitly constructed under some regularity conditions on W.

6.5 Construction of convergent rational spectral densities with converging canonical spectral factors

In this section we give the main ideas for the construction of a sequence of rational spectral densities with CSFs converging to the true CSF. Let $\{W_r\}_{r\in\mathbb{N}}$ be a sequence of rational spectral densities having no poles on $\mathbb{T}$. We define

$$c_k = \frac{1}{2\pi}\int_{\mathbb{T}} W(z)z^{-k}\mu(dz) \quad \text{and} \quad c_{k,r} = \frac{1}{2\pi}\int_{\mathbb{T}} W_r(z)z^{-k}\mu(dz) \quad k = 0, 1, \ldots$$

The sequences $\{c_k\}_{k\in\mathbb{N}}$ and $\{c_{k,r}\}_{k\in\mathbb{N}}$ are the unique covariance sequences associated with W and W_r, respectively. By the Riemann-Lebesque Lemma, $c_k \to 0$ as $k \to \infty$. The covariance sequence $c_{k,r}$ has the form:

$$c_{k,r} = C_r A_r^k B_r + \sum_{m=0}^{m_r} D_{m,r}\Delta(k-m), \tag{6.4}$$

where A_r, B_r, C_r, and $D_{0,r}, D_{1,r}, \ldots, D_{m_r,r}$ are $n \times n$ matrices with A_r having eigenvalues in $\mathbb{D}$, (A_r, B_r, C_r) is a minimal realization, and $\Delta(m)=\begin{cases} 1 & \text{if m=0} \\ 0 & \text{otherwise} \end{cases}$. The central idea of our construction is to require the sequence $\{W_r\}_{r\in\mathbb{N}}$ to satisfy

$$\deg \Phi(W_r) \leq nd_r, \tag{6.5a}$$

$$c_{k,r} = c_k \quad \text{for} \quad k = 0, 1, \ldots, d_r, \tag{6.5b}$$

where $\{d_r\}_{r\in\mathbb{N}}$ is an increasing sequence of positive integers. That a sequence $\{W_r\}_{r\in\mathbb{N}}$ satisfying (6.5) exists and is computable is the content of the theory of rational covariance extension with degree constraint [9, 16, 6, 48, 17, 18]. Since $\|W - W_r\|_1 \geq \sup_{k\geq 0} \|c_k - c_{k,r}\|_1 \geq \sup_{0\leq k\leq d_r} \|c_k - c_{k,r}\|_1$, we see that the discrepancy between the first few terms of the covariance sequence of W and W_r yields a lower bound for the approximation error in $\mathcal{L}^1_{n\times n}$. Therefore, it makes sense to impose the condition (6.5b). Moreover, since it is desirable to have W_r be as "simple" as possible, the constraint (6.5a) is also well-motivated. Plugging in the Fourier series expansion of W_r in the definition of $\|W - W_r\|_1$, we obtain:

$$\begin{aligned}
&\int_{\mathbb{T}} \|W(z) - W_r(z)\|_1 \mu(dz) \\
&\quad \leq \int_{\mathbb{T}} \left\| W(z) - \sum_{k=0}^{d_r} \Re\{c_k z^k\} \right\|_1 \mu(dz) + \left\| \int_{\mathbb{T}} \Re\Big\{ C_r A_r^{d_r+1} (I - A_r z)^{-1} B_r \right. \\
&\qquad \left. + I_{\{d_r \leq m_r - 1\}}(d_r) \sum_{m=d_r+1}^{m_r} D_{m,r} z^m \Big\} \mu(dz) \right\|_1, \\
&\quad \leq \int_{\mathbb{T}} \left\| W(z) - \sum_{k=0}^{d_r} \Re\{c_k z^k\} \right\|_1 \mu(dz) + \mathrm{R}(W_r, d_r),
\end{aligned} \tag{6.6}$$

where $I_A(x)$ is the indicator function for the set A and

$$\mathrm{R}(W_r, d_r) = \left\| \Re\Big\{ C_r A_r^{d_r+1} \int_{\mathbb{T}} (I - A_r z)^{-1} \mu(dz) B_r \Big\} \right\|_1.$$

If W satisfies $\operatorname{ess\,sup}_{z\in\mathbb{T}} \|W(z)\|_1 < \infty$, the Fourier series of W converges to W in $\mathcal{L}^2_{n\times n}$, hence also in $\mathcal{L}^1_{n\times n}$. Therefore, the first term on the right hand side of (6.6) goes to 0 as $r \to \infty$ and the following theorem is immediate:

Theorem 6.5.1 *Suppose* $\operatorname{ess\,sup}_{z\in\mathbb{T}} \|W(z)\|_1 < \infty$ *and let* $\{W_r\}_{r\in\mathbb{N}}$ *be a sequence of rational spectral densities satisfying* Assumption A4, A5 *and the interpolation constraints of (6.5). If* $\lim_{r\to\infty} \mathrm{R}(W_r, d_r) = 0$ *then* Assumption A2 *holds and* $\lim_{r\to\infty} \|\Phi(W) - \Phi(W_r)\|_2 = 0$.

It is reasonable to expect, at least intuitively, that there could be "many" sequences which satisfy the condition of Theorem 5.1 if the spectral density W is not too "irregular" . Indeed, we see later in Corollary 6.5.3 and Corollary 6.5.5 some particular instances where this is true. Moreover, we will show that the approximating sequence $\{W_r\}$ can be constructed explicitly under some further, yet mild, assumptions on W.

6.5.1 The scalar case

We shall give a constructive proof of the following result:

Theorem 6.5.2 *Let W be a continuous scalar spectral density and $\|W-U/V\|_\infty = 0$, where U and V are, respectively, continuous and Lipschitz spectral densities. If $\{U_r\}_{r\geq 1}$ is a sequence of non-negative definite pseudopolynomials converging uniformly to U then the following statements hold:*

1. *If $V > 0$, there is a sequence $\{V_r\}_{r\geq 1}$ of non-negative pseudopolynomials such that $W_r = U_r(V_r)^{-1}$ satisfies (6.5b) (with $n = 1$) for all r sufficiently large*

2. *If $V \not> 0$ but i) U (hence also V) is zero only at a finite number of points on $\mathbb{T}$, ii) $W(e^{i\theta}) = 0$ whenever $V(e^{i\theta}) = 0$, iii) $U_r(e^{i\theta}) = 0$ only if $U(e^{i\theta}) = 0$ and iv) $\sup_r \|U_r/U\|_\infty < \infty$, there is a sequence $\{V_r\}_{r\geq 1}$ of non-negative pseudopolynomials such that $W_r = U_r(V_r)^{-1}$ satisfies $\frac{1}{2\pi}\int_{\mathbb{T}} W_r(z)z^{-k}\mu(dz) = c_k - s_{k,r}$ where $s_{k,r} = \sum_{l=1}^{k_r} K_l e^{ik\theta_l}$, $k_r \in \mathbb{N}$ is at most $d_r - 1$, $K_1, \ldots, K_{k_r}$ are some non-negative constants, and $\theta_1, \ldots, \theta_{k_r} \in (-\pi, \pi]$ with $\theta_i \neq \theta_j$ if $i \neq j$*

In either case, $\{V_r\}_{r\geq 1}$ and $\{W_r\}_{r\geq 1}$ converge in $\mathcal{L}^\infty$ to V and W, respectively. Moreover, if $\deg U_r \leq 2d_r$ then W_r also satisfies (6.5a).

Let l^2 denote the (standard) set of all square-summable infinite sequences. Let Γ_c denote the real linear space of all real-valued continuous functions on $\mathbb{T}$. Define the real linear space $\mathcal{F}_c$ as

$$\mathcal{F}_c = \{(q_0, q_1, \ldots) \in l^2 \mid q_k = \tfrac{1}{2\pi}\textstyle\int_{-\pi}^{\pi} e^{-ik\theta}W(e^{i\theta})d\theta \ \forall k \geq 0, \text{ for some } W \in \Gamma_c\}.$$

Since a continuous function f is uniquely determined by its Fourier coefficients [78, Theorem 2.4] and since the negative Fourier coefficients are merely conjugates of the positive Fourier coefficients whenever f is a real-valued function, $\mathcal{F}_c$ is actually isomorphic to Γ_c. Therefore, we may uniquely identify any element of $\mathcal{F}_c$ with an element of Γ_c, and vice-versa. By endowing Γ_c with a topology induced by the supremum norm and endowing $\mathcal{F}_c$ with a topology induced by the norm $\|q\| = \operatorname*{ess\,sup}_{\theta\in(-\pi,\pi]}\left|\sum_{k=0}^{\infty} \Re\{q_k e^{ik\theta}\}\right|$ (since any element of Γ_c is continuous, the infinite sum converges pointwisely for almost all $z \in \mathbb{T}$ [78, Chapter 19]), we in fact have a homeomorphism from $\mathcal{F}_c$ to Γ_c. Moreover, $\mathcal{F}_c$ is then a closed set since Γ_c is. In

the sequel, we denote the homeomorphic map from $\mathcal{F}_c$ to Γ_c by Q. The map Q is linear: $Q(a_1 f_1 + a_2 f_2) = a_1 Q(f_1) + a_2 Q(f_2)$ for any $a_1, a_2 \in \mathbb{R}$ and any $f_1, f_2 \in \mathcal{F}_c$. We now define some relevant convex subsets of Γ_c and $\mathcal{F}_c$. Define Γ_c^+ to be the convex cone consisting of elements of Γ_c which are non-negative on $\mathbb{T}$. We define the convex cone $\mathcal{F}_c^+$ analogously to $\mathcal{F}_c$ by replacing Γ_c with Γ_c^+. In a similar manner, we see that $\mathcal{F}_c^+$ is isomorphic to Γ_c^+. Endowing Γ_c^+ (resp. $\mathcal{F}_c^+$) with a topology derived from Γ_c (resp. $\mathcal{F}_c$), we also get that $\mathcal{F}_c^+$ is homeomorphic to Γ_c^+ under Q, i.e., if Q^+ is the restriction of Q to $\mathcal{F}_c^+$ then Q^+ is a homeomorphism from $\mathcal{F}_c^+$ to Γ_c^+. Define D_r to be the subset of $\mathcal{F}_c^+$ consisting of all $q = (q_0, q_1, \ldots) \in \mathcal{F}_c^+$ such that $q_k = 0 \ \forall k > r$ and the trigonometric polynomial $\sum_{k=0}^{r} \Re\{q_k e^{ik\theta}\} \geq 0$ $\forall \theta \in (-\pi, \pi]$. Clearly, $D_s \supset D_r$ if $s > r$. Moreover, since the partial Fejér sums of any $f \in \Gamma_c^+$ are non-negative pseudopolynomials and approximate f arbitrarily closely (for details on Fejér sums see [78]), it is immediate that $\overline{\cup_{r\geq 0} D_r} = \mathcal{F}_c^+$.

Recall that a function f on $\mathbb{T}$ is *Lipschitz* if $\|f(e^{i\theta}) - f(e^{i\psi})\|_1 \leq K|\theta - \psi|$ $\forall \theta, \psi \in (-\pi, \pi]$ for some positive constant K, and observe that a scalar spectral density $W \in \Gamma_c^+$ can be written as $W = U/V$ a.e., where V is any Lipschitz scalar spectral density and $U = WV$. Let W have the covariance sequence $\mathbf{c} = \{c_0, c_1, \ldots\} \in l^2$. Define $\mathbf{c}_r = \mathrm{col}(c_0, c_1, \ldots, c_r)$ to be the partial covariance sequence of $\mathbf{c}$ up to the r^{th} term, and let $\mathbf{d}_r$ be an arbitrary element of D_r. The functional $\mathbb{J}^r : D_r \to \mathbb{R} \cup \{\infty\}$, parametrized by $\mathbf{c}_r$ and $\mathbf{d}_r$, is defined as:

$$\mathbb{J}^r(q; \mathbf{c}_r, \mathbf{d}_r) = \Re\{\sum_{k=0}^{r} \min\{k+1, 2\} c_k^* q_k - \langle Q_r(\mathbf{d}_r), \log Q_r(q) \rangle\}, \tag{6.7}$$

where $\langle f, g \rangle = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(e^{i\theta}) g_*(e^{i\theta}) d\theta$ and Q_r is a map with domain D_r defined by:

$$Q_r(q_0, q_1, \ldots, q_r, 0, 0, \ldots)(e^{i\theta}) = \sum_{k=0}^{r} \Re\{q_k e^{ik\theta}\}$$

Notice that Q_r can be viewed as the restriction of Q to D_r. The functional $\mathbb{J}^r$ has been introduced and analyzed in chapters 3 and 4 (actually, our formulation here is slightly different from the previous chapters. However, it causes no difficulty since the functionals in those chapters can be recovered by application of the linear invertible transformation $(q_0, q_1, \ldots, q_r) \mapsto (q_0, \frac{1}{2} q_1, \ldots, \frac{1}{2} q_r)$ and redefining D_r in an obvious way). $\mathbb{J}^r$ has the following properties:

P1. $\mathbb{J}^r$ is strictly convex on D_r and is continuous at all points except the origin.

P2. $\mathbb{J}^r$ has compact sub-level sets and a unique minimizer in D_r.

P3. $\mathbb{J}^r$ is infinitely differentiable along any line lying in the interior of D_r.

Moreover, if $\mathbb{J}^r$ has a minimizer q_s which is stationary (i.e., the gradient is zero at q_s) then $\frac{Q(d_r)}{Q(q_s)}$ satisfies $\langle \frac{Q(d_r)}{Q(q_s)}, g_k \rangle = c_k$ for $k = 0, 1, \ldots, r$, where $g_k(z) = z^k$. Note that q_s is always a stationary point whenever $Q(\mathbf{d}_r)$ is positive definite [6, Theorem 4.10].

Let us now consider another functional $\mathbb{J} : \mathcal{F}_c^+ \to \mathbb{R} \cup \{\infty\}$ parametrized by the covariance sequence $\mathbf{c}$ and an infinite sequence $\mathbf{d} \in \mathcal{F}_c^+$ such that $Q(\mathbf{d}) = U$. It is defined as:

$$\mathbb{J}(q; \mathbf{c}, \mathbf{d}) = \Re\{\mathrm{H}(q; \mathbf{c}) - \langle Q(\mathbf{d}), \log Q(q) \rangle\} \tag{6.8}$$

where $\mathrm{H}(\cdot; \mathbf{c})$ is a linear function on $\mathcal{F}_c$, parametrized by $\mathbf{c}$, defined by:

$$\mathrm{H}(q; \mathbf{c}) = \lim_{r \to \infty} \sum_{k=0}^{r} \min\{k+1, 2\} c_k^* q_k. \tag{6.9}$$

Since $q \in l^2$ whenever $q \in \mathcal{F}_c$ (recall the definition of $\mathcal{F}_c$) and $\mathbf{c} \in l^2$, it follows that $\sum_{k=0}^{\infty} \min\{k+1, 2\} c_k^* q_k < \infty$. Therefore, $\mathrm{H}(\cdot; \mathbf{c})$ is well defined $\forall q \in \mathcal{F}_c$. Let us define the convex set $D_e(\mathbb{J}) = \{q \in \mathcal{F}_c^+ \mid \int_{\mathbb{T}} \log^- Q(q)(z) \mu(dz) < \infty\}$ (where $\log^- x = \max\{0, -\log x\}$); $D_e(\mathbb{J})$ is actually the *effective domain* (see, for example, [79]) of $\mathbb{J}$. Then clearly $\cup_{r \geq 1} D_r \subset D_e(\mathbb{J})$. Since $\|\cdot\|$ is also a norm on $D_e(\mathbb{J})$, we endow $D_e(\mathbb{J})$ with the topology induced by the $\|\cdot\|$ norm (this is precisely the relative topology of $D_e(\mathbb{J})$ as a subset of $\mathcal{F}_c^+$: open sets in $D_e(\mathbb{J})$ are sets of the form $D_e(\mathbb{J}) \cap O$ for any O which is an open set of $\mathcal{F}_c^+$). Continuing on, along the same line of arguments as for $\mathbb{J}^r$ we may verify that $\mathbb{J}$ has property P4 (given below), and property P3 with $\mathbb{J}^r$ and D_r replaced by $\mathbb{J}$ and $D_e(\mathbb{J})$, respectively (an analogue of property P2 need not hold for $\mathbb{J}$ and will not be required in the following).

P4. $\mathbb{J}$ is strictly convex on $D_e(\mathbb{J})$ and continuous on the interior of $D_e(\mathbb{J})$.

In the remaining analysis, let us view $\mathbb{J}_\Psi$ as a convex functional that maps from the convex set $D_e(\mathbb{J})$ to $\mathbb{R}$. We now derive an expression for the directional derivatives of $\mathbb{J}$ following [17, 18]. Define

$$\mathcal{M}_{\mathbf{d}} = \{q \in D_e(\mathbb{J}) \mid \operatorname*{ess\,sup}_{\theta \in (-\pi, \pi]} Q(\mathbf{d})(e^{i\theta})(Q(q)(e^{i\theta}))^{-1} < \infty\}.$$

Let $q' \in \mathcal{F}_c^+$, $0 < h < 1$, and suppose that $q \in \mathcal{M}_{\mathbf{d}}$. We observe that if $Q(q + h(q' - q))(z) = (1-h)Q(q)(z) + hQ(q')(z_0) = 0$ for some $z_0 \in \mathbb{T}$ and all

$0 < h < 1$, then $Q(q)$ and $Q(q')$ must share a zero at the point z_0. On the other hand, if $Q(q+h(q'-q))(z) > 0$ for all $z \in \mathbb{T}$ then $Q(q)$ and $Q(q')$ cannot possibly have a zero in common on $\mathbb{T}$. As a result, by the mean-value theorem of calculus, we obtain:

$$\begin{aligned} & Q(\mathbf{d})(z)^{\frac{\log Q(q+h(q'-q))(z)-\log Q(q)(z)}{h}} \\ &= Q(\mathbf{d})(z)^{\frac{\partial}{\partial v}\log Q(q+v(q'-q))(z)\Big|_{v=\eta(h,z)}} \\ &= Q(\mathbf{d})(z)^{\frac{Q(q')-Q(q)}{Q(q)(z)+\eta(h,z)Q(q'-q)(z)}} \end{aligned} \tag{6.10}$$

for all $z \in \mathbb{T}$ such that $Q(q)(z) > 0$ (hence for almost all $z \in \mathbb{T}$ since $q \in \mathcal{M}_{\mathbf{d}}$), where $0 < \eta(h,z) \leq h$. Moreover, $q \in \mathcal{M}_{\mathbf{d}}$ implies that

$$\operatorname*{ess\,sup}_{(h,z)\in[0,\frac{1}{2}]\times\mathbb{T}} \frac{Q(\mathbf{d})(z)}{Q(q)(z)+\eta(h,z)Q(q'-q)(z)} < \infty. \tag{6.11}$$

Now, the directional derivative $\nabla_{q'-q}\mathbb{J}$ at q in the direction $q'-q$ is defined as

$$\nabla_{q'-q}\mathbb{J}(q;\mathbf{c},\mathbf{d}) = \lim_{h\downarrow 0}\frac{\mathbb{J}(q+h(q'-q);\mathbf{c},\mathbf{d})-\mathbb{J}(q;\mathbf{c},\mathbf{d})}{h}. \tag{6.12}$$

Plugging in the definition of $\mathbb{J}$ into (6.12), and using (6.10) and (6.11) with the Lebesque Dominated Convergence Theorem to bring h under the integral (this is essentially the same argument used in [17]), and finally evaluating the limit as $h \downarrow 0$, we obtain: $\nabla_{q'-q}\mathbb{J}(q;\mathbf{c},\mathbf{d}) = H(q'-q) - \langle \frac{\Psi}{Q(q)}, Q(q') - Q(q)\rangle$. Let $\mathcal{F}^+_{c,\infty}$ denote the set of all elements a in $D_e(\mathbb{J})$ for which $Q(a)(e^{i\theta})$ is infinitely differentiable with respect to θ, and suppose that $Q(q)$ is Lipschitz and $Q(q')$ in $\mathcal{F}^+_{c,\infty}$. Then $\sum_{k=0}^{r} q_k e^{ik\theta}$ (resp. $\sum_{k=0}^{r} q'_k e^{ik\theta}$) converges uniformly to $Q(q)$ (resp. $Q(q')$) [80, Theorem 2, p. 142]. By plugging in the definition of H and by another application of the Lebesque Dominated Convergence Theorem, we get:

$$\nabla_{q'-q}\mathbb{J}(q;\mathbf{c},\mathbf{d}) = \lim_{r\to\infty}\sum_{k=0}^{r}\min\{k+1,2\}\Re\left\{\left(c_k^* - \int_{\mathbb{T}}\frac{Q(\mathbf{d})(z)}{Q(q)(z)}z^k\mu(dz)\right)(q'_k - q_k)\right\},$$

for all $q \in \mathcal{M}_{\mathbf{d}}$ and all $q' \in \mathcal{F}^+_{c,\infty}$. However, since $\mathcal{F}^+_{c,\infty}$ is dense in $\mathcal{F}^+_c$, hence also in $D_e(\mathbb{J})$, the preceding expression for $\nabla_{q'-q}\mathbb{J}(q;\mathbf{c},\mathbf{d})$ is valid for all $q \in \mathcal{M}_{\mathbf{d}}$ and all $q' \in \mathcal{F}^+_c$. Now, setting $q_\# = Q^{-1}(V)$, hence $Q(q_\#)$ is Lipschitz and $q_\# \in \mathcal{M}_{\mathbf{d}}$, we obtain $\sum_{k=0}^{r}\min\{k+1,2\}\left(c_k - \int_{\mathbb{T}}\frac{Q(\mathbf{d})(z)}{Q(q_\#)(z)}z^{-k}\mu(dz)\right)^* = 0$ for all r. It therefore

follows that $\nabla_{q'-q_\#}\mathbb{J}(q_\#;\mathbf{c},\mathbf{d}) = 0\ \forall q' \in \mathcal{F}_c^+$, so $q_\#$ is a stationary point of $\mathbb{J}$ and, by Property P4, must also be the unique minimizer of $\mathbb{J}$.

Suppose now that $\{\mathbf{d}_r\}_{r\in\mathbb{N}}$ is such that $\lim_{r\to\infty}\|\mathbf{d}_r - \mathbf{d}\| = 0$ with $\mathbf{d} = Q^{-1}(U)$ as defined previously. Let $\mathbb{J}\big|_{D_r}\colon D_r \to \mathbb{R}\cup\{\infty\}$ denote the restriction of $\mathbb{J}$ to D_r defined by $\mathbb{J}\big|_{D_r}(q;\mathbf{c},\mathbf{d}) = \Re\Big\{\sum_{k=0}^{r}\min\{k+1,2\}c_k^* q_k - \langle Q(\mathbf{d}), \log Q_r(q)\rangle\Big\}$. Since $Q(d) = U$ is continuous on $\mathbb{T}$ and is zero only on a subset of $\mathbb{T}$ of μ-measure zero, the analysis of $\mathbb{J}^r$ in [5, 6] and chapters 3 and 4 readily carries over to $\mathbb{J}\,|_{D_r}$ to show that the latter also has properties P1, P2, and P3 (with $\mathbb{J}^r$ replaced by $\mathbb{J}\big|_{D_r}$). Moreover, defining $s_r = \arg\min_{q\in D_r}\mathbb{J}\big|_{D_r}(q;\mathbf{c},\mathbf{d})$, then it is clear that $\lim_{r\to\infty} s_r = q_\#$. Hence, for any $\epsilon > 0$ we will have $\|q_\# - s_r\| < \frac{\epsilon}{2}$ by taking large enough r. Then also $\|q - q_\#\| \le \|q - s_r\| + \|s_r - q_\#\| \le \epsilon$ whenever $\|q - s_r\| \le \frac{\epsilon}{2}$ (recall that $q, q_\#, s_r \in D_e(\mathbb{J})$). Therefore, for all r large enough, $\{q \in D_r \mid \|q - s_r\| \le \frac{\epsilon}{2}\} \subset \{q \in D_e(\mathbb{J}) \mid \|q - q_\#\| \le \epsilon\}$ and:

$$\begin{aligned}
\sup_{\{q\in D_r, \|q-s_r\|\le\frac{\epsilon}{2}\}} |\mathbb{J}\big|_{D_r} - \mathbb{J}^r| &= \sup_{\{q\in D_r | \|q-s_r\|\le\frac{\epsilon}{2}\}} |\langle Q(\mathbf{d}) - Q(\mathbf{d}_r), \log Q_r(q)\rangle| \\
&\le \sup_{\{q\in D_e(\mathbb{J}) | \|q-q_\#\|\le\epsilon\}} |\langle Q(\mathbf{d}) - Q(\mathbf{d}_r), \log Q(q)\rangle| \\
&\le \|Q(\mathbf{d}) - Q(\mathbf{d}_r)\| D,
\end{aligned}$$

where $D = \max_{\{q\in D_e(\mathbb{J}) | \|q-q_\#\|\le\epsilon\}}\left|\int_{-\pi}^{\pi}\log Q(q)(e^{i\theta})d\theta\right|$. Since $\lim_{r\to\infty}\|Q(\mathbf{d}) - Q(\mathbf{d}_r)\| = 0$, it then follows that $\lim_{r\to\infty}\max_{\{q\in D_r | \|q-s_r\|\le\frac{\epsilon}{2}\}}\left|\mathbb{J}\big|_{D_r}(q;\mathbf{c},\mathbf{d}) - \mathbb{J}^r(q;\mathbf{c}_r,\mathbf{d}_r)\right| = 0$ for any $\epsilon > 0$. Due to properties P1, P2, and P3 of $\mathbb{J}\big|_{D_r}$ and $\mathbb{J}^r$, the preceding limit implies that $\lim_{r\to\infty}\|s_r - q_{\#,r}\| = 0$, where $q_{\#,r} = \arg\min_{q\in D_r}\mathbb{J}^r(q;\mathbf{c}_r,\mathbf{d}_r)$; for the details refer to the proof of Theorem 3.3.8 in Chapter 3 (replace Ψ, Ψ_k, $\mathbb{J}_\Psi$ and $\mathbb{J}_{\Psi_k}$ with $Q(\mathbf{d})$, $Q(\mathbf{d}_k)$, $\mathbb{J}\big|_{D_k}$ and $\mathbb{J}^k$, respectively). Furthermore, since $\|q_\# - q_{\#,r}\| \le \|q_\# - s_r\| + \|s_r - q_{\#,r}\|$ and both terms on the right tend to 0 as $r\to\infty$, we get $\lim_{r\to\infty}\|q_\# - q_{\#,r}\| = 0$ and $\lim_{r\to\infty}\|Q(q_\#) - Q(q_{\#,r})\| = 0$. Consequently, if $V > 0$ then $q_\#$ is in the interior of $D_e(\mathbb{J})$ and the same is true for $q_{\#,r}$ for all r sufficiently large. Therefore, W_r satisfies (6.5) for all r sufficiently large, and $W_r = \frac{Q(\mathbf{d}_r)}{Q(q_{\#,r})} \to W = \frac{Q(\mathbf{d})}{Q(q_\#)}$ uniformly as $r\to\infty$.

If $V \not> 0$ but assumptions i-iv in Point 2 of the theorem are satisfied, convergence of $\{W_r\}$ to W in $\mathcal{L}^\infty$ can again be established. Let $V_r = Q(q_{\#,r})$. Since $V_r \to V$ uniformly, we have that for any $\delta > 0$ $\exists R(\delta) \in \mathbb{N}$ such that $-\delta < V_r(e^{i\theta}) - V(e^{i\theta}) < \delta\ \forall\theta$ and $\forall r > R(\delta)$. However, since $V, V_r \ge 0$

and V, V_r are continuous $\forall r$, and $\exists R' \in \mathbb{N}$ such that $V_r(e^{i\theta}) > 0 \ \forall r > R'$ whenever $V(e^{i\theta}) > 0$ (due to assumptions i and iii and the observation that all zeros of V_r on $\mathbb{T}$ are also zeros of U_r [17, 18] and $V_r \to V$ uniformly), there is a *continuous* function $f_\delta : \mathbb{T} \to [0,1]$ such that a) $V - \delta f_\delta \geq 0$, b) $(V - \delta f_\delta)(e^{i\theta}) = 0$ if and only if $V(e^{i\theta}) = 0$, and c) $-\delta f_\delta < V_r - V < \delta \ \forall \theta$ and $\forall r > \max\{R(\delta), R'\}$. In particular, letting $S(\delta) = \max\{R(\delta), R'\}$ we may always take f_δ to be $f_\delta = \sup_{r>S(\delta)}(V - V_r)/\delta$ (and we shall do so in the sequel) and satisfy all the requirements. Then $V/V_r < V/(V - \delta f_\delta) = 1/(1 - \delta f_\delta/V)$ a.e. $\mathbb{T} \ \forall r > S(\delta)$. Moreover, $1/(1 - g_\delta/V) \in \mathcal{L}^\infty$ for all δ sufficiently small, where $g_\delta = \delta f_\delta = \sup_{r>S(\delta)}(V - V_r)$. To see this, first observe that property a of f_δ implies $g_\delta/V \leq 1$ a.e. $\mathbb{T}$. Then we observe that $g_\delta \downarrow 0$ (i.e., g converges monotonically to 0) uniformly on $\mathbb{T}$ as $\delta \downarrow 0$ and hence, since also $g_\delta \leq V \ \forall \delta > 0$, for sufficiently small δ we will have $g_\delta < V$ for all $z \in \mathbb{T}$ except those for which $V(z) = 0$. Therefore, $\|g_\delta/V\|_\infty < 1$ and $1/(1 - g_\delta/V) < \infty$ a.e. for sufficiently small δ, as claimed, and it follows that $\sup_{r>S(\delta)} \|V/V_r\|_\infty < \infty$. Now, let $N_{\delta'}(z) = \{y \in \mathbb{T} \mid |z - y| < \delta'\}$ for any $z \in \mathbb{T}$ and $\delta' > 0$. Then we note that V/V_r (resp. $V/V_r - 1$) converges uniformly to 1 (resp. 0) on $\mathbb{T} \backslash \cup_{k=1}^m N_{\delta'}(z_k)$, where $z_1, \ldots, z_m \in \mathbb{T}$ are all zeros of V, for δ' small enough such that $\cup_{k=1}^m N_{\delta'}(z_k)$ is a strict subset of $\mathbb{T}$. By uniform convergence of U_r to U and assumption iv, an analogous remark is also true for U_r/U. Next, we make the observation that $|W_r - W| = |U_r/V_r - U/V| \leq W|U_r/U - 1|V/V_r + W|V/V_r - 1|$. Then the properties of V/V_r and U_r/U just stated, along with assumptions i-ii and the continuity of W, imply that both $W|U_r/U - 1|V/V_r$ and $W|V/V_r - 1|$ converge uniformly to 0 as $r \to \infty$. To see this, let us consider the term $W|U_r/U - 1|V/V_r$ and let R'' be large enough such that $M = \sup_{r>R''} \|U_r/U - 1\|_\infty \|V/V_r\|_\infty < \infty$ (recall that $\sup_{r>S(\delta)} \|V/V_r\|_\infty < \infty$). Then, by assumption ii and the continuity of W, for any $\epsilon > 0$ we may choose $\delta' > 0$ small enough such that $\sup_{\{z \in \cup_{k=1}^m N_{\delta'}(z_k)\}} W(z) < \epsilon/M$ a.e. followed by choosing $r > R''$ which is large enough such that $\sup_{\{z \in \mathbb{T} \backslash \cup_{k=1}^m N_{\delta'}(z_k)\}} |U_r(z)/U(z) - 1|V(z)/V_r(z) < \epsilon/\|W\|_\infty$. In other words, for any $\epsilon > 0 \ \exists R'''(\epsilon)$ such that $\|W|U_r/U - 1|V/V_r\|_\infty < \epsilon \ \forall r > R'''(\epsilon)$. The same line of arguments may then be applied to $W|V/V_r - 1|$. In conclusion, we again have $W_r \overset{\mathcal{L}^\infty}{\to} W$ as $r \to \infty$. That $\frac{1}{2\pi} \int_{\mathbb{T}} W_r(z) z^{-k} \mu(dz) = c_k - s_{k,r}$ as stated in the theorem has been shown in the proof of [18, Theorem 8].

Finally, we note that the preceding analysis remains valid if $\mathbf{d}_r \in D_r$ is replaced by $\mathbf{d}_r \in \cup_{k>r} D_k$. To see this, let $\mathbf{d}_r$ be any pseudopolynomial, *not necessarily* of degree at most r, and define $\mathbb{J}^r$ as in (6.7). Then it may be verified that

$\mathbb{J}^r$ again has properties P1, P2, and P3, and if q_s is a stationary point minimizer of $\mathbb{J}^r$ then $\frac{Q(d_r)}{Q(q_s)}$ once more satisfies $\langle \frac{Q(d_r)}{Q(q_s)}, g_k \rangle = c_k$ for $k = 0, 1, \ldots, r$. The rest of the analysis follows *mutatis mutandis*. This completes the proof of Theorem 6.5.2.

An important consequence of the theorem combined with Corollary 6.4.6 and [19, Theorem 2] is the following:

Corollary 6.5.3 *Suppose $W \in \mathcal{L}^1$ is a continuous spectral density with a finite number of zeros on $\mathbb{T}$ and let $\{W_r\}_{r\geq 1}$ be a sequence as defined in Theorem 6.5.2. Then $\lim_{r\to\infty} \|\Phi(W_r) - \Phi(W)\|_2 = 0$. If, in addition, $W > 0$ and $\frac{d}{d\theta}W(e^{i\theta}) \in \mathcal{L}^2$ then also $\lim_{r\to\infty} \|\Phi(W_r) - \Phi(W)\|_\infty = 0$.*

Notice that the corollary gives a weaker condition for convergence in $\|\cdot\|_\infty$ norm than analyticity (resp. rationality and boundedness) and positivity of W given in [20, Theorem 1] (resp. [13, Theorem 3.4]) for the Szegö-Levinson algorithm, and does not restrict $\Phi(W_r)$ to have all its zeros at the origin. Note that we say W is analytic if it can be continued analytically from all points in $\mathbb{T}$.

6.5.2 The matrix case

For a matrix-valued spectral density W, the situation is slightly more complicated. If W is a matrix-valued Lipschitz spectral density then we may write $W = (W^{-1})^{-1} = \det(W)\mathrm{adj}(W)^{-1}$, where $\mathrm{adj}(W)$ denotes the adjoint of W. Define $U = P \det W$ and $V = P\,\mathrm{adj}(W)$ for any arbitrary scalar spectral density P which is Lipschitz and positive definite. Then U and V are Lipschitz. The representation $W = UV^{-1}$ a.e. can be viewed as the matricial counterpart of the scalar fractional representation. If W is positive definite then so is V and in this case, by suitably redefining the sets Γ_c, $\mathcal{F}_c$, Γ_c^+, $\mathcal{F}_c^+$, D_r and the associated norms with their respective matricial counterparts, as well as suitably modifying the functionals $\mathbb{J}^r$ and $\mathbb{J}$ (see [55, eq. (V.5), p. 2180]), it is a relatively straightforward, but tedious, exercise to adapt the analysis developed in deriving Theorem 6.5.2 to the matrix case. Then we may show the following counterpart of Theorem 6.5.2:

Theorem 6.5.4 *Let $W = UV^{-1} \in \mathcal{L}^1_{n\times n}$ be a matrix-valued positive definite Lipschitz spectral density, where $U = P \det W$ and $V = P\mathrm{adj}(W)$ for some positive definite Lipschitz scalar spectral density P. If $\{U_r\}_{r\geq 1}$ is a sequence of positive*

definite pseudopolynomials converging uniformly to U then there exists a (unique) sequence $\{V_r\}_{r\geq 1}$ of positive definite pseudopolynomials such that:

1. *$W_r = U_r(V_r)^{-1}$ satisfies (6.5b) for all r. If, in addition, $\deg U_r \leq 2nd_r$, then (6.5a) is also satisfied.*

2. *$\{V_r\}_{r\geq 1}$ and $\{W_r\}_{r\geq 1}$ converge uniformly to V and W, respectively.*

It then follows from Corollary 6.4.6 and [19, Theorem 2]:

Corollary 6.5.5 *Let $\{W_r\}_{r\in\mathbb{N}}$ be a sequence as defined in Theorem 6.5.4. Then $\lim_{r\to\infty} \|\Phi(W_r) - \Phi(W)\|_2 = 0$. If, in addition, $W > 0$ and $\frac{d}{d\theta}W(e^{i\theta}) \in \mathcal{L}^2_{n\times n}$ then also $\lim\limits_{r\to\infty} \|\Phi(W_r) - \Phi(W)\|_\infty = 0$.*

It is plausible that Theorem 6.5.4 and Corollary 6.5.5 can be extended to the case where U has zeros on $\mathbb{T}$. However, to do this, we must allow some spectral zeros (see [55]) of W_r to be on $\mathbb{T}$. This is currently an open problem.

6.6 A spectral factorization algorithm

We now introduce a new algorithm for spectral factorization of a special class $\mathcal{W}_n$ of spectral densities. $\mathcal{W}_1$ denotes the set of spectral densities $W \in \mathcal{L}^1$ which can be continued analytically from every point $z \in \mathbb{T}$ except from a finite number of points $w_k = e^{i\theta_k}$, $k = 1, \ldots, M$, for which $W(w_k) = 0$ and $\lim_{z\in\mathbb{T}, z\to w_k} \frac{|z-w_k|^{m_k}}{W(z)} < \infty$ for some integer $m_k \geq 1$. For $n > 1$, $\mathcal{W}_n$ denotes the set of spectral densities in $\mathcal{L}^1_{n\times n}$ which are positive definite and can be continued analytically from every point on $\mathbb{T}$. We state the algorithm below followed by a discussion of the steps involved and a convergence analysis.

6.6.1 The algorithm

Given: A spectral density $W \in \mathcal{W}_n$, the desired accuracy $\epsilon > 0$ and maximum number of iterations $r_{\max}$.

Initialize: Normalize W so that $c_0 = I$. Let $e^{i\lambda_1}, \ldots, e^{i\lambda_L} \in \mathbb{T}$ be local minima of $\det W$ satisfying $0 \leq \det W(e^{i\lambda_l}) \leq \alpha$ ($\alpha \in \mathbb{R}$, $\alpha \geq 0.2$. Rule of thumb: $\alpha = 0.2$). Let $\mathcal{V}_1$ be all points in $\{e^{i\lambda_1}, \ldots, e^{i\lambda_L}\}$ from which $\det W$ does not have an analytic continuation and $\mathcal{V}_2 = \{e^{i\lambda_1}, \ldots, e^{i\lambda_L}\}\backslash\mathcal{V}_1$. For $l = 1, \ldots, L$, define $m_l = \min\{k \in \mathbb{N} \mid \lim_{z\in\mathbb{T}, z\to e^{i\lambda_l}} \frac{|z-e^{i\lambda_l}|^k}{W(z)} < \infty\}$ if

$e^{i\lambda_l} \in \mathcal{V}_1$ and $m_l = \min\{k \in \mathbb{N} \mid D_\theta^k \det W(e^{i\lambda_l}) \neq 0\}$ if $e^{i\lambda_l} \in \mathcal{V}_2$ (here $D_\theta^m \det W(e^{i\lambda_l}) = \frac{d^m \det W(e^{i\theta})}{d\theta^m}\big|_{\theta=\lambda_l}$). Let $\eta_0(z) = \prod_{l=1}^{L}(z - v_l)^{\frac{m_l}{2}}$, where $v_l = \max\{0, r_l\}e^{i\lambda_l}$ (with $r_l = 1 - \left(\frac{\det W(e^{i\lambda_l})}{D_\theta^{m_l} \det W(e^{i\lambda_l})}\right)^{\frac{1}{m_l}}$) if $e^{i\lambda_l} \in \mathcal{V}_2$, and $v_l = e^{i\lambda_l}$ if $e^{i\lambda_l} \in \mathcal{V}_1$. Set $r = 1$, $d_0 = L$, and compute $c_0, c_1, \ldots, c_L$ and the outer polynomial matrix $R_0 = \phi(V_0)$, where $V_0 = Q(\arg\min_{q \in D_{d_0}} \mathbb{J}^{d_0}(q))$ (see Section 5).

Step 1. Select a point $z_r \in \mathbb{D}$. Then:

(a) If W is symmetric (i.e., $W(e^{-i\theta}) = W(e^{i\theta})$) or $\theta_r \notin \{0, \pi\}$, set $d_r = d_{r-1} + 2$, and $\eta_r = \eta_{r-1}(z - z_r)(z - z_r^*)$, otherwise

(b) Set $d_r = d_{r-1} + 1$, and $\eta_r = \eta_{r-1}(z - z_r)$.

Step 2. Compute $c_{d_{r-1}+1}, \ldots, c_{d_r}$ and the outer matrix polynomial $R_r = \phi(V_r)$, where $V_r = Q(\arg\min_{q \in D_{d_r}} \mathbb{J}^{d_r}(q))$.

Step 3. Compute $e_r = \frac{1}{2}\|W - W_r\|_1 + \frac{1}{2}\|\Phi(W_r) - \Phi(W_{r-1})\|_2$, where $W_r = \eta_{r*}\eta_r(R_r R_{r*})^{-1}$. If $e_r > \epsilon$ and $r \leq r_{\max}$, set $r = r + 1$ and return to Step 1.

End: $z^{d_r}\eta_{r*}(R_r)^{-1}\sqrt{c_0}$ is the approximate CSF.

Computation of the polynomial matrix R_r, $r = 0, 1, 2, \ldots$, is given in [50, 55] and Chapter 5 of the thesis. The main idea of the algorithm is to find a sequence $z_1, z_2, \ldots \in \overline{\mathbb{D}}$ such that $W_r = U_r V_r^{-1}$ satisfies (6.5) and $W_r \to W$ in $\mathcal{L}_{n\times n}^{\infty}$, where $U_r(z) = \eta_0\eta_{0*}\Pi_{k=1}^{d_r-L}(z - z_k)(z - z_k)_*$. It works as follows. Suppose $e^{i\lambda_l} \in \mathcal{V}_2$, then $\det W$ has an analytic continuation to some open set containing $e^{i\lambda_l}$. Moreover, if $D_\theta^m \det W(e^{i\lambda}) = 0$ for $m = 1, \ldots, l$ then also $(\det W)^{(m)}(e^{i\theta}) = (\det W)^{(m)}(z)\big|_{z=e^{i\lambda_l}} = 0$ ($(\det W)^{(m)}$ denotes the m^{th} derivative of the analytic continuation of $\det W$). Since $e^{i\lambda_l}$ is a local minimum, we have that $D_\theta^{m_l} \det W(e^{i\lambda_l}) > 0$. Let us take care of points $z \in \mathbb{T}$ for which $\det W(z) \approx 0$. We take these to be the points $e^{i\lambda_1}, \ldots, e^{i\lambda_L}$ as defined in the algorithm. For $e^{i\lambda_l} \in \mathcal{V}_2$, the Taylor series expansion of $\det W(z)$ about $e^{i\lambda_l}$ gives $\det W(z) \approx \det W(e^{i\lambda_l}) + (-ie^{-i\theta})^{m_l} D_\theta^{m_l} \det W(e^{i\lambda_l})(z - e^{i\lambda_l})^{m_l}$ for z sufficiently close to $e^{i\lambda_l}$. To estimate a zero of $\det W(z)$ about $e^{i\lambda_l}$, we set $\det W(z) = 0$ to get $|z - e^{i\lambda_l}| \approx \left(\frac{\det W(e^{i\lambda_l})}{D_\theta^{m_l} \det W(e^{i\lambda_l})}\right)^{\frac{1}{m_l}}$. Assuming $z_l = r_l e^{i\lambda_l}$ with $0 \leq r_l \leq 1$ for our zero estimate, we obtain $|1 - r_l| = \left(\frac{\det W(e^{i\lambda_l})}{D_\theta^{m_l} \det W(e^{i\lambda_l})}\right)^{\frac{1}{m_l}}$. Thus, we choose

$r_l = 1 - \left(\frac{\det W(e^{i\lambda_l})}{D_\theta^{m_l} \det W(e^{i\lambda_l})}\right)^{\frac{1}{m_l}}$ and set $v_l = \max\{0, r_l\}e^{i\lambda_l}$ (hence automatically $v_l = e^{i\lambda_l}$ if $\det W(e^{i\lambda_l}) = 0$). As elaborated in Section 1, points $z \in \mathbb{T}$ for which $\det W(z) \approx 0$ slows convergence down significantly due to slow decay of the so-called Schur parameters [9]. The main idea in the algorithm is to reduce their influence by suitably placing a zero of η_0 in their vicinity as in [9, 5, 6], but here we allow the degree of the approximation W_r to increase as required.

Remark 6.6.1 *$v_1, \ldots, v_L$ in $\mathbb{D}$ actually serve as estimates of zeros of $\det W$ in some open annulus $\{z \in \mathbb{C} \mid 1-\delta < |z| < 1\}$ $(0 < \delta < 1)$. As such, other schemes can be used to determine these points. The "rule of thumb" $\alpha = 0.2$ is based on the subjective view that it is "not too small" and "not too large". If convergence of the algorithm is slow, say, $e_r > 10^{-2}$ in the first few (5-10) iterations, one may try restarting the algorithm with α increased, or the next remark may be taken into consideration.*

Remark 6.6.2 *If $\det W$ has thin and sharp "spectral line"-like peaks then the algorithm may perform poorly. This is because such a peak indicates the possible presence of a (non-cancelling) pole and zero close to each other and to the unit circle, while the zero is not included in η_0 [9]. To remedy the situation, let $H \in \mathcal{H}^2$ be a scalar notch filter with narrow stop bands around frequencies corresponding to the peaks, $P = H_*H$ and apply the algorithm to $W' = WP$. Then $\Phi(W) \approx \frac{\Phi(W')}{H}$ in $\mathcal{H}^2$.*

The following theorem gives a requirement on $z_1, z_2, \ldots$ for convergence:

Theorem 6.6.3 *Let η_r be as defined in the algorithm. Suppose that the polynomial $\rho_r(z) = z^{d_r}\frac{\eta_{r*}(z)}{\eta_{0*}(z)} \overset{\mathcal{L}^\infty}{\to} \rho$, where ρ is continuous and has no zeros on $\mathbb{T}$. Let $W \in \mathcal{W}_1$ (resp. $\mathcal{W}_n$, $n > 1$), $U_r(z) = \eta_{r*}(z)\eta_r(z)$ and V_r is as defined in Theorem 6.5.2 (resp. Theorem 6.5.4). Then $z^{d_r}\eta_{r*}(\Phi(V_r))^{-1}$ converges to $\Phi(W)$ in $\mathcal{H}^2$ (resp. $\mathcal{H}^2_{n\times n}$), and also in $\mathcal{H}^\infty$ (resp. $\mathcal{H}^\infty_{n\times n}$) if $W > 0$ and, when $n = 1$, $\frac{d}{d\theta}W(e^{i\theta}) \in \mathcal{L}^2$.*

Proof. Assume that W has been normalized so that $c_0 = I$. Let $U = \eta_0\rho_*\rho\eta_{0*}$ and define V by $V(e^{i\theta}) = \lim_{\lambda\to\theta} U_*(e^{i\lambda})U(e^{i\lambda})W(e^{i\lambda})^{-1}$. Then, by definition, $V \in \mathcal{L}^\infty_{n\times n}$ and $\|W - UV^{-1}\|_\infty = 0$. Note that $\eta_r = z^{d_r}\eta_0\rho_{r*}$ and let $U_r = \eta_{r*}\eta_r$. Then, since $\rho_r \overset{\mathcal{L}^\infty}{\to} \rho$, we have that $U_r \overset{\mathcal{L}^\infty}{\to} U$. By Theorem 6.5.2 and Corollary 6.5.3, or Theorem 6.5.4 and Corollary 6.5.5, whichever pair is applicable, it follows that V_r and $W_r = U_rV_r^{-1}$ converge in $\mathcal{L}^\infty_{n\times n}$, respectively, to V and W, and

$\Phi(W_r)$ converges to $\Phi(W)$ in $\mathcal{H}^2_{n\times n}$ and also in $\mathcal{H}^\infty_{n\times n}$ if $W > 0$ and, when $n = 1$, $\frac{d}{d\theta}W(e^{i\theta}) \in \mathcal{L}^2$ (by the definition of $\mathcal{W}_n$, $\frac{d}{d\theta}W(e^{i\theta}) \in \mathcal{L}^2_{n\times n}$ is automatically satisfied when $W \in \mathcal{W}_n$, $n > 1$). Since $\Phi(W_r) = \Phi(U_r)(\Phi(V_r))^{-1} = z^{d_r}\eta_{r*}(\Phi(V_r))^{-1}$, scaling back by multiplication of both $\Phi(W)$ and $\Phi(W_r)$ on the right with $\sqrt{c_0}$ gives the desired result. □

Remark 6.6.4 *Clearly, if $z_r = 0$ for all $r > R$ ($R \in \mathbb{N}$) then ρ_r converges uniformly to the analytic function $\rho = \prod_{k=1}^{R}(1 - z_k^* z)$ and the algorithm converges.*

6.6.2 General approximation strategy

The algorithm requires $W \in \mathcal{W}_n$ for some $n \in \mathbb{N}$. If this is not the case but W is continuous and has a finite number of zeros, then the strategy would be to first construct an approximating analytic spectral density (which need not be rational) in $\mathcal{L}^\infty_{n\times n}$. Then we apply the spectral factorization algorithm to the approximation to obtain an approximate CSF of $\Phi(W)$. The fact that the analytic approximation does not have to be rational affords us flexibility in choosing a set of basis functions for the approximation.

6.6.3 Heuristic scheme for selection of spectral zeros

In Theorem 6.6.3 we gave an explicit condition on the spectral zeros $z_1, z_2, \ldots$ for the spectral factorization algorithm to converge and mention a particular situation where this condition is automatically met. In the following we give an intuitive heuristic scheme for choosing $z_1, z_2, \ldots$ for scalar W. The idea goes as follows. For each r (including $r = 0$) we have at Step 1 that $\int_{-\pi}^{\pi}\left(W_r(e^{i\theta}) - W(e^{i\theta})\right)d\theta = 0$. If $W_r - W$ is not identically zero (for which the algorithm then terminates), then it is easy to show, using the mean value theorem of calculus, that $\exists\theta$ such that $W_r(e^{i\theta}) - W(e^{i\theta}) > 0$. Since a zero of W_r can decrease the magnitude of W_r in certain regions of $\mathbb{T}$, the main idea now is to try to reduce the excess (or overshoot) of W_r over W at a point θ_r for which the excess is relatively large (preferably the largest). If W is not symmetric or $\theta_r \in \{0, \pi\}$ then we place a zero at $z_r = Re^{i\theta_r}$ (with $0 < R < 1$ so that $z_r \in \mathbb{D}$) such that $\frac{W_{r-1}(z)|z-z_r|^2}{W(z)}\Big|_{z=e^{i\theta_r}} = \frac{W_{r-1}(e^{i\theta_r})}{W(e^{i\theta_r})}(1 - R)^2 = 1$. From the last equality we obtain the required value of R for Step 1. In case W is symmetric and $\theta_r \notin \{0, \pi\}$, we must place two zeros at z_r and z_r^* to ensure W_r is also symmetric. By

a procedure similar to the symmetric case, we find that a quartic equation $|1-R|^2|1-Re^{-i2\theta_r}|^2 - \frac{W_{r-1}(e^{i\theta_r})}{W(e^{i\theta_r})} = 0$ must be solved for R and a real solution satisfying $0 < R < 1$ is chosen. It is easy to see, since $\frac{W_{r-1}(e^{i\theta_r})}{W(e^{i\theta_r})} < 1$, that the quartic solution always has such a solution. It is not theoretically guaranteed that spectral zeros chosen by the scheme satisfies the requirements of Theorem 6.6.3 for convergence. However, in accordance with Remark 6.6.4, we may always proceed with the heuristic for a *finite* number of steps before terminating the selection by setting $z_r = 0$ for the remaining iterations. Simulation results to be given in Section 6.7, however, indicate that this heuristic seems to work reasonably well.

6.6.4 Reduction of computational time

The computationally intensive part in the proposed algorithm is Step 2 for computing R_r. This is because the homotopy continuation algorithms described in [50, 55] involve solving a finite sequence of convex optimization problems. However, it is important to note that the computation can be substantially reduced at higher iterations down to solving only *one* convex optimization problem. To see this, consider the case where the algorithm is convergent and $E_r = \|z^{d_r}\eta_{r*}R_r^{-1} - z^{d_{r-1}}(\eta_{r-1})_*R_{r-1}^{-1}\|_\infty \to 0$ as $r \to \infty$. Since R_r is invertible a.e. $\mathbb{T}$ and $\sup_{r\geq 1}\|R_r\|_\infty < \infty$, we also have $E_r' = \|z^{d_r}\eta_{r*}R_{r-1} - z^{d_{r-1}}(\eta_{r-1})_*R_r\|_\infty \to 0$ as $r \to \infty$ (by noting $E_r' \leq \|R_{r-1}\|_\infty E_r \|R_r\|_\infty$). Assuming for the moment that $z_r \in \mathbb{R}$ and recalling that $\eta_r(z) = (z - z_r)\eta_{r-1}(z)$ and $d_r = d_{r-1} + 1$, we get that $\|(1 - z_r^* z)R_{r-1} - R_r\|_\infty \to 0$. Therefore, $\|(1 - z_r^* z)R_{r-1} - R_r\|_\infty$ will be small for all r sufficiently large. In that case, we simply set the homotopy step-size parameter ρ (resp. λ) in [50] (resp. [55]) to 1 and use the coefficients of $(1 - z_r^* z)R_{r-1}$ as an initial point in the algorithm for solving the *single* convex optimization problem which gives the coefficients of R_r. If $z_r \notin \mathbb{R}$, then replace $(1 - z_r^* z)$ with $(1 - z_r z)(1 - z_r^* z)$. This reduction scheme can be executed when $e_r \leq \delta$ for some small $\delta > 0$.

If required, further reduction is possible. We note that the Hessian of the functional to be minimized has a Hankel-plus-Toeplitz structure which can be inverted (or solved if it is the coefficient matrix in a system of linear equations) with fast algorithms given in [56, 57]. More importantly, however, is that these algorithms have *parallel* (i.e., the Schur-type) versions which can be implemented on parallel computers.

6.7 Numerical examples

In this section we apply the new spectral factorization algorithm and heuristic of the last section to compute approximate CSF's of some rational and non-rational non-coercive spectral densities. In each example, three different simulations are carried out:

1. Simulation A: The spectral zero selection heuristic is applied at Step 1 until termination of the algorithm.

2. Simulation B: The spectral zero selection heuristic is applied at Step 1 for a finite, pre-specified, number of steps after which z_r is set to 0.

3. Simulation C: All of Initialize are skipped except the computation of c_0, z_r is set to 0 in Step 1 for all r, and R_r in Step 2 is computed recursively via the Szegö-Levinson algorithm. Step 3 is unaltered.

We set $\epsilon = 10^{-4}$ in all simulations and apply the computational reduction scheme of Section 6.6.4 in Simulation A and B when $e_r \leq 10^{-2}$ is satisfied. The algorithm was implemented in Matlab and executed on a computer with a Pentium 4 processor with a clock speed of 3.2 GHz and 1 GB of RAM.

Example 6.7.1 *Consider the rational spectral density* $W(e^{i\theta}) = \frac{2+\cos\theta-2\cos 2\theta}{24.1-18.9\cos\theta+2\cos 2\theta}$ *which is non-coercive with a zero at* $z=-1$. *The exact CSF of* W *is known to be* $\Phi(W)(z) = -\sqrt{10}\frac{(z-2)(z+1)}{(z-4)(z-5)}$. *The results of Simulation A, B, and C are shown in Table 6.1 and Fig. 6.1 (in simulation B, only 5 zeros are selected with the heuristic scheme). The exact error* $\|\Phi(W_{10}) - \Phi(W)\|_2$ *for Simulation A was* 4.74526×10^{-5}.

Table 6.1: Simulation results

	Number of iterations	Final value of e_r	Degree of approximation	Running time (seconds)
Simulation A	10	4.12909×10^{-5}	10	8.53
Simulation B	10	4.14906×10^{-5}	10	7.593
Simulation C	151	1.98136×10^{-3}	151	73.063

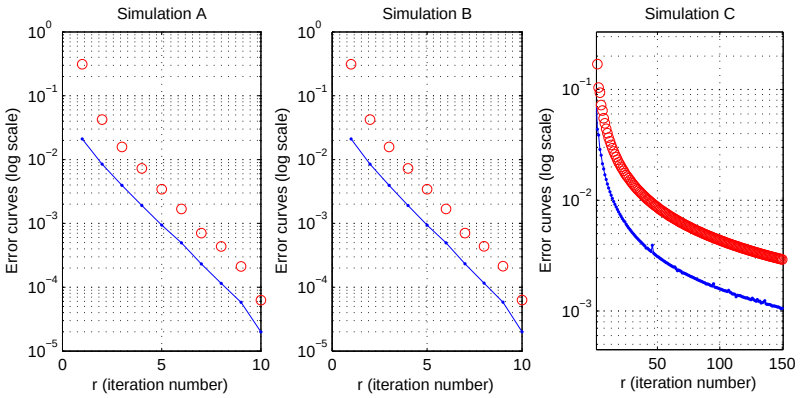

Figure 6.1: Plots of Simulation A (left), B (center), C (right): $\|W - W_r\|_1$ (dash-dot line), $\|\Phi(W_r) - \Phi(W_{r-1})\|_2$ (circle)

Example 6.7.2 *The Kolmogorov spectral density [81], which is the spectral density of a continuous time stochastic process arising in the study of turbulence, is defined along the imaginary axis as* $W_K(i\omega;\sigma) = \frac{1}{\sqrt{1-\sigma(i\omega)^2}}$ *where* σ *is a positive parameter. To use our approach, we first transform the spectral density from the imaginary axis to the unit circle via the (invertible) bilinear transformation* $e^{i\theta} = \frac{1-i\omega}{1+i\omega}$. *After applying the transformation we get a spectral density* W_K^d *on* $\mathbb{T}$ *given by:* $W_K^d(e^{i\theta};\sigma) = \sqrt{\frac{1+\cos\theta}{1+\cos\theta+\sigma(1-\cos\theta)}}$. *Notice that* W_K^d *has a zero at* $z = -1$ *which is not PLL (see discussions on PLL in Sections 2.6 and 2.7) and cannot be continued analytically from that point. Setting* $\sigma = 2$, *results from Simulation A, B and C are shown in Table 6.2 and Fig. 6.2 (in Simulation B, only 10 zeros are selected with the heuristic scheme).*

Table 6.2: Simulation results

	Number of iterations	Final value of e_r	Degree of approximation	Running time (seconds)
Simulation A	22	9.32203×10^{-5}	43	239.269
Simulation B	29	9.31331×10^{-5}	39	294.741
Simulation C	151	2.11347×10^{-3}	151	44.86

Transforming the approximate CSF of Simulation A from the unit circle back to the imaginary axis gives us the frequency response shown in Fig. 6.3.

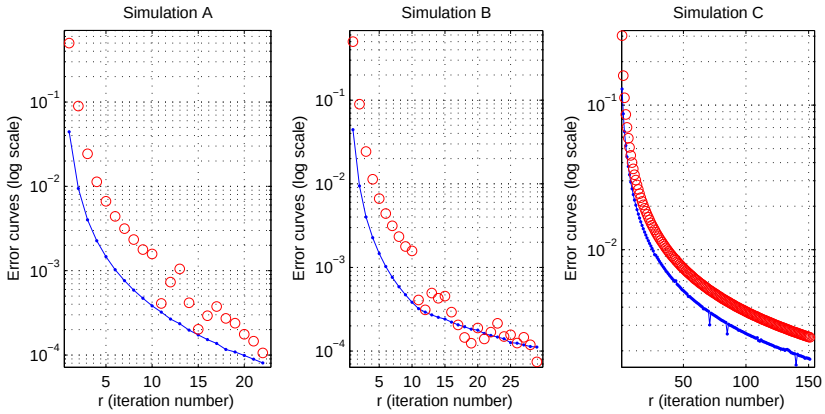

Figure 6.2: Plots of Simulation A (left), B (center), C (right): $\|W - W_r\|_1$ (dash-dot line), $\|\Phi(W_r) - \Phi(W_{r-1})\|_2$ (circle)

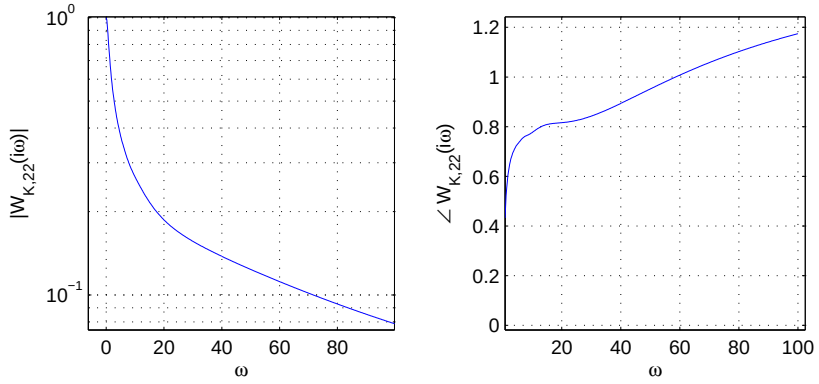

Figure 6.3: Frequency response of approximate CSF from Simulation A

Example 6.7.3 *The* von Karman *spectral density [30, p. 73] is the spectral density of a continuous time stochastic process defined along the imaginary axis as* $W_{vK}(i\omega;\sigma) = 2\sigma \frac{1-\frac{8}{3}\sigma^2(1.339)^2(i\omega)^2}{(1-\sigma^2(1.339)^2(i\omega)^2)^{\frac{11}{6}}}$*, where* σ *is a positive parameter. It is often used as a substitute for the Kolmogorov power spectral density of the previous example. After a transformation from the real line to the unit circle, we obtain a spectral density* W^d_{vK} *on* $\mathbb{T}$ *given by:*

$$W^d_{vK}(e^{i\theta};\sigma) = 2\sigma \frac{1+\cos\theta+\frac{8}{3}\sigma^2(1.339)^2(1-\cos\theta)}{(1+\cos\theta+\sigma^2(1.339)^2(1-\cos\theta))^{\frac{11}{6}}}(1+\cos\theta)^{\frac{5}{6}}\,.$$

W^d_{vK} *has a zero at* $z=-1$ *which, as with the previous example, is not PLL and cannot be continued analytically from that point. Setting* $\sigma = 2$*, the results of Simulation A, B and C are shown in Table 6.3 and Fig. 6.4 (in Simulation B, only 10 zeros are selected with the heuristic scheme).*

Table 6.3: Simulation results

	Number of iterations	Final value of e_r	Degree of approximation	Running time (seconds)
Simulation A	30	6.54517×10^{-5}	49	926.656
Simulation B	47	8.97805×10^{-5}	47	1014.2
Simulation C	151	7.98211×10^{-3}	151	64.392

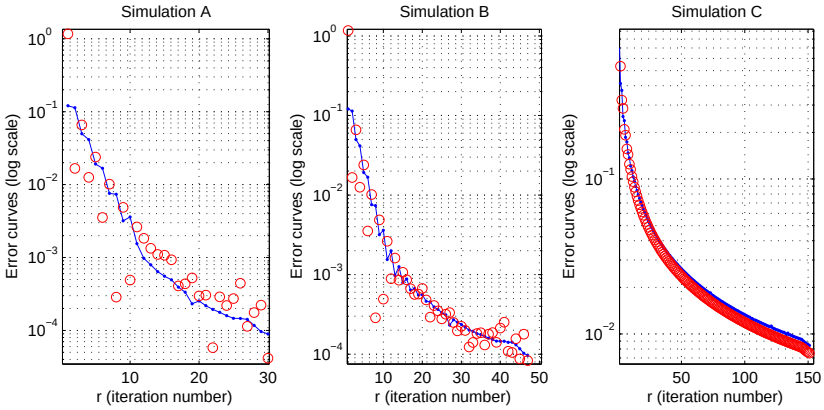

Figure 6.4: Plots of Simulation A (left), B (center), C (right): $\|W - W_r\|_1$ (dash-dot line), $\|\Phi(W_r) - \Phi(W_{r-1})\|_2$ (circle)

Transforming the approximate CSF of Simulation A from the unit circle to the imaginary axis gives us the frequency response as shown in Fig. 6.5.

All examples indicate that both Simulation A and B give better results than Simulation C (the Szegö-Levinson algorithm). Despite producing an approximation of substantially higher order, Simulation C gives a final error e_r of magnitude 10^2 higher and it would seem many hundred more iterations are required to achieve $e_r < 10^{-4}$ as in Simulation A and B. In Example 7.1 for a simple second order spectral factor, Simulation C also runs much longer. Simulation A runs faster than B, but gives an approximation of slightly higher degree. The latter is not unexpected since Simulation B selects more real-valued zeros (i.e., at the origin). The simulations do suggest that the heuristic is practically useful and quite effective, regardless of whether a limited or indefinite number of spectral zeros are selected.

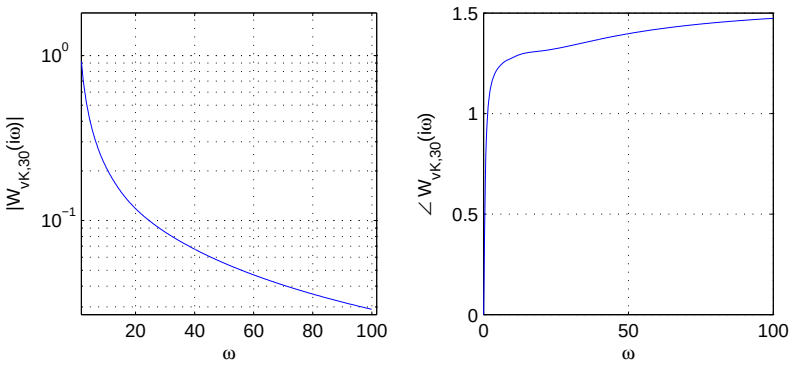

Figure 6.5: Frequency response of approximate CSF from Simulation A

6.8 Conclusions and further research

This chapter makes three primary contributions. First and foremost, we have derived a set of sufficient, easy to verify conditions for uniform log-integrability of a sequence of matrix-valued spectral densities. Secondly, we establish theoretical results on the existence of certain approximating rational sequences for a class of matrix-valued spectral densities. Finally, we propose a new spectral factorization algorithm for a more specific class of matrix-valued spectral densities based on degree constrained rational covariance extensions, and establish convergence results. Our approach does not require the spectral density to be coercive. There is a freedom to choose a sequence of spectral zeros in the algorithm and a heuristic has been proposed for choosing them. The performance of the new algorithm is demonstrated in a number of numerical examples, where it performed favorably compared to the popular Szegö-Levinson algorithm/maximum entropy method. In particular, the algorithm was successfully applied to the non-rational and non-coercive Kolmogorov and von Karman spectral densities. Possible topics for future research include development of fast algorithms for computing degree constrained covariance extensions (as discussed in Section 4.4), relaxations of the conditions presented here, and development of better heuristics for selection of spectral zeros.

The results and algorithm of the chapter may be useful in applications in which spectral factorization plays a prominent role such as in computation of approximate solutions of algebraic Riccati equations (ARE's) in optimal control of linear systems, or in which signals with non-rational power spectra is a central theme (e.g., control of aircraft subject to windgust, adaptive optics, and laser

scintillation [30]). It may also prove to be useful in spectral estimation and system identification research.

Part II

Topics in Quantum Linear Stochastic Systems

List of Notation and Terminology for Part II

Notation

$\mathbb{R}$	The set of real numbers
$\mathbb{C}$	The set of complex numbers
^{T}	Transpose of a matrix/array
$[\cdot,\cdot]$	For two Hilbert space operators A, B which maps a Hilbert space into itself, $[A, B]$ denotes the commutator of A, B defined by: $[A, B] = AB - BA$, assuming that the products AB and BA are well-defined on a common dense domain if either A or B or both are unbounded
$\otimes$	Tensor product
*	The adjoint of a (possibly unbounded) Hilbert space operator
$^\#$	If X is a matrix/array of Hilbert space operators , $X^\#$ denotes the operation of taking the adjoint of each element of X
†	$X^\dagger = (X^\#)^T$
$\mathrm{diag}(\cdot,\ldots,\cdot)$	For square matrices $M_1, \ldots, M_n$, $\mathrm{diag}(M_2, \ldots, M_n)$ denotes a block diagonal matrix with matrices $M_1, \ldots, M_n$ on the diagonal block
$\mathrm{diag}_m(\cdot)$	For a square matrix T, $\mathrm{diag}_m(T)$ denotes the block diagonal matrix $\mathrm{diag}(T, \ldots, T)$ where T appears m times as a diagonal block

$I_{n\times n}$	Denotes the $n \times n$ identity matrix. If n is not specified, it is assumed that it can be determined from the context
$0_{n\times m}$	Denotes the $n\times m$ zero matrix. If n and/or m are not specified, it is assumed that they can be determined from the context
P_m	A $2m \times 2m$ permutation matrix defined by $$P_m a = [\ a_1 \quad a_3 \quad \dots \quad a_{2m-1} \quad a_2 \quad a_4 \quad \dots \quad a_{2m}\ ]^T,$$ where $a = [\ a_1 \quad a_2 \quad \dots \quad a_{2m}\ ]^T$ and $a_1, \dots, a_{2m} \in \mathbb{C}$

Terminology

Permutation matrix	A full-rank real matrix whose columns (or, equivalently, rows) consist of standard basis vectors for $\mathbb{R}^m$; i.e., vectors in $\mathbb{R}^m$ whose elements are all 0 except for one element which has the value 1. A permutation matrix P has the unitary property $PP^T = P^T P = I$
Hilbert space operator	An operator mapping from one Hilbert space to another
Unitary operator	A bounded Hilbert space operator, say U, possessing the unitary property $U^*U = UU^* = I$, where I is the identity operator
Commute	Two vectors x, y of operators on a common Hilbert space are said to commute if $$xy^T - (yx^T)^T = 0$$ on a dense subspace of the Hilbert space

Chapter 7

Quantum Linear Stochastic Systems in Quantum Optics

7.1 Introduction

Recent successes in quantum and nano-technology have provided a great impetus for research in the area of quantum feedback control systems; e.g., see [82, 83, 84, 85, 86, 87, 88]. It is reasonable to expect that quantum control is an area of research which could play a vital role towards realization of conceptual quantum signal processing systems and quantum computers that are being extensively studied for potential benefits over their classical counterparts [89].

One particular area in which significant theoretical and experimental advances have been achieved is *quantum optics*. In particular, *linear quantum optics* is one of the possible platforms being investigated for building future quantum computers [90], [89, Section 7.5], besides being an area of independent interest for physicists. It is especially interesting from the point of view of an engineer, and of a control theorist in particular, for two primary reasons:

1. A prominent mathematical tool for modelling of quantum optical devices is quantum stochastic calculus, developed by Hudson and Parthasarathy [91, 92], which is a generalization to the quantum context of the classical Ito stochastic calculus. The latter is of course a familiar tool widely used by engineers in filtering and estimation of stochastic dynamical systems.

2. Under the so-called rotating wave approximation and weak coupling assumption (see detailed discussions in Chapters 3 and 5 of [93]), the approx-

imate dynamics of various devices in quantum optics takes on a form similar to that of a linear time invariant stochastic system of modern systems theory. More precisely, these devices are conveniently modelled as *quantum linear stochastic systems* represented by a set of *quantum linear stochastic differential equations* (QSDEs) driven by non-commutative/quantum Wiener process:

$$\begin{aligned} dx(t) &= Ax(t)dt + Bdw(t); \\ dy(t) &= Cx(t)dt + Ddw(t), \end{aligned} \tag{7.1}$$

with A, B, C, D being constant matrices. Here w denotes a vector of non-commutative Wiener processes. Regarding $w(t) = [\ w_1(t) \quad \dots \quad w_{n_w}(t)\ ]^T$, each operator-valued stochastic process $w_i(t)$, $i \in \{1, \dots, n_w\}$, when considered independently is equivalent to a classical Wiener process on some classical probability space via the Spectral Theorem [94, Theorem 2.4]. The distinction with classical Wiener processes is that two processes $w_i(t)$ and $w_j(t)$, $i \neq j$, need not commute. That is, $w_i(s)w_j(t) - w_j(t)w_i(s)$ need not be zero for any $s, t \geq 0$. In this case, a joint distribution cannot be prescribed for the two processes on the *same* classical probability space. This means that they cannot be measured simultaneously, a distinctive feature of quantum mechanics.

Due to the above observations, it comes naturally to ask whether one can generalize various controller synthesis paradigms from modern control theory, such as the LQG and H^∞ paradigms, to the quantum optical domain. To make a case for this possibility, recently James and Petersen [32] have proposed a generalization of the H^∞ synthesis method to quantum linear systems based on a quantum extension of the Strict Bounded Real Lemma [95] of the classical theory. For a given disturbance attenuation level, if a controller exists then it can be synthesized via solving a pair of Riccati equations, similar to the classical case. However, the synthesis only gives a *partial model* of the controller, i.e, not all of the system matrices are fully prescribed. In an example given in [32], where the plant to be controlled is a (quantum) optical cavity, they demonstrated that the controller partial model can be completed by appropriately adding some quantum Wiener noise and realized as another optical cavity. Thus, one interesting question which arises from the H^∞ synthesis procedure which we attempt to address in this thesis is:

"Is it always possible to add quantum noises to complete a partial model in such a way that the controller models some physically meaningful quantum linear system, such as a quantum optical cavity, linear amplifier or attenuator?"

In general, quantum linear stochastic systems represented by linear QSDEs with arbitrary constant coefficients need not correspond to some physically meaningful system. This is unlike classical linear stochastic systems (throughout this part of the thesis we shall use the term "classical" to loosely refer to systems which have no quantum mechanical components), such as those considered in Part I of this thesis, which may be regarded as always being realizable, at least approximately, via electronics and/or mechanical devices. Physical quantum systems must satisfy some additional constraints which impose some algebraic conditions on the system coefficients A, B, C, D. One such constraint is that physical systems must preserve the *canonical commutation relations* (CCR) among certain *canonical quantum observables*. For example, in a basic one dimensional quantum harmonic oscillator [96] set on the Hilbert space $L^2(\mathbb{R})$, the space of measurable and square integrable complex-valued functions on $\mathbb{R}$, the canonical observables are the position operator $p : f(x) \mapsto xf(x)$ and momentum operator $q : f(x) \mapsto \frac{d}{dx}f(x)$, or the annihilation operator $a = q + ip$ and creation operator $a^* = p - iq$ (here * denotes the adjoint of an operator), defined on an appropriate dense subspace of $L^2(\mathbb{R})$, and the CCR takes the form (in the Schrödinger picture):

$$[p, q] = i,$$

or equivalently,

$$[a, a^*] = 1,$$

where we take the Planck constant $\hbar$ to be 1. In the Heisenberg picture [96] the CCR takes the form:

$$[p(t), q(t)] = i \quad \forall t \geq 0,$$

or equivalently,

$$[a(t), a(t)^*] = 1 \quad \forall t \geq 0,$$

where $k(t)$, k can be either p, q, a or a^*, are the time evolution of k under a unitary evolution generated by the operator-valued Hamiltonian H of the quantum harmonic oscillator given by $H = \frac{1}{2}(p^2 + q^2)$, i.e., $k(t) = U_t^\dagger k U_t$ and $U_t = e^{iHt}$, $\forall t \geq 0$.

Note that for convenience some authors may adopt a different convention regarding the definition of position and momentum operators p and q than the

one given above, e.g. [97, 98]. Sometimes it is defined as scalar multiple of the operator p and q as defined above such as $p : f(x) \mapsto \sqrt{2}xf(x)$ and $q : f(x) \mapsto \sqrt{2}\frac{d}{dx}f(x)$, and the annihilation and creation operators as $a = \frac{p+iq}{2}$ and $a^* = \frac{p-iq}{2}$. Hence, the CCR in this case becomes $[p, q] = 2i$ and $[a, a^*] = 1$. This causes no difficulties as long as the definitions and the CCR are used consistently, and is more a matter of preference. In this chapter and the next we shall adhere to the latter convention (see a footnote in Section 7.2).

We first give a description of quantum linear stochastic systems which will be the focus of our investigation. Due to the substantial amount of background materials required to set up Part II, we shall omit them and assume that the reader has some familiarity with the principles of quantum mechanics, quantum probability spaces, quantum stochastic processes, and quantum stochastic calculus. An introduction to quantum mechanics on finite dimensional Hilbert spaces which only requires knowledge of elementary linear algebra, but which shows most of the essential features of quantum mechanics, can be found in [89, Chapter 2]. For a more general treatment of quantum mechanics, the reader may refer to the standard text [96]. For an introduction to quantum probability spaces suitable for engineers with working knowledge of applied functional analysis, the reader may consult the tutorial paper [94], while for an introduction to quantum stochastic processes (which includes quantum Wiener processes) and quantum stochastic calculus, the reader may refer to the original paper of Hudson and Parthasarathy [91], the text [92], Chapter 5 of [99], or the tutorial paper [94].

This chapter is an adaptation of the paper [100] (joint work with M. R. James and I. R. Petersen). The contributions of the chapter include the derivation of a necessary and sufficient condition for preservation of the CCR in quantum linear stochastic systems (Theorem 7.3.1), the introduction of a formal notion of physical realizability of linear quantum stochastic systems (Definition 7.3.5), and explicit necessary and sufficient condition for physical realizability of such systems (Theorem 7.3.6).

7.2 General quantum linear stochastic models for quantum optics

We are generally interested in physical systems that contain one or more components that are quantum in nature. It is helpful to have in mind an interconnection

of components, some of which are "classical", meaning that non-quantum descriptions suffice, and some for which "quantum" descriptions are required. Such systems are common in quantum optics laboratories, and may occur, for instance, in schemes for implementing quantum computing and information processing algorithms. We use non-commutative or quantum probability theory to describe the systems of interest. This framework is quite general and encompasses quantum and classical mechanical systems. Quantum noise, which may arise from measurements or interactions between subsystems and the environment, plays a central role.

To be specific, the systems we consider can be defined on some quantum probability space $(\mathscr{A}, \mathbb{P})$ (e.g., see [94] and the references therein), where $\mathscr{A}$ is a von Neumann algebra (of bounded operators on some Hilbert space) and $\mathbb{P}$ is a state on this algebra. The von Neumann algebra can be thought of as an abstract mathematical representation of the "observables" or physical quantities of interest, while statistical attributes of these observables are determined by $\mathbb{P}$. We begin by presenting a general form of the system of interest, followed by a formal discussion of the associated Hilbert spaces and algebras.

We consider linear non-commutative stochastic systems of the form

$$\begin{aligned} dx(t) &= Ax(t)dt + Bdw(t); \quad x(0) = x_0 \\ dy(t) &= Cx(t)dt + Ddw(t) \end{aligned} \tag{7.2}$$

where A, B, C and D are, respectively, real $\mathbb{R}^{n\times n}, \mathbb{R}^{n\times n_w}, \mathbb{R}^{n_y\times n}$ and $\mathbb{R}^{n_y\times n_w}$ matrices (n, n_w, n_y are positive integers), and $x(t) = [\ x_1(t) \ \ \ldots \ \ x_n(t)\ ]^T$ is a vector of self-adjoint possibly non-commutative system variables.

The initial system variables $x(0) = x_0$ consist of operators (on an appropriate Hilbert space) satisfying the *commutation relations**

$$[x_j(0), x_k(0)] = 2i\Theta_{jk}, \quad j,k = 1,\ldots,n, \tag{7.3}$$

where Θ is a real antisymmetric matrix with components Θ_{jk}, and $i = \sqrt{-1}$. Here, the commutator is defined by $[A, B] = AB - BA$. To simplify matters without loss of generality, we take the matrix Θ to be of one of the following forms:

*In the case of a single degree of freedom quantum particle, $x = (x_1, x_2)^T$ where $x_1 = q$ is the position operator, and $x_2 = p$ is the momentum operator. The annihilation operator is $a = (q + ip)/2$. The commutation relations are $[a, a^*] = 1$, or $[q, p] = 2i$.

- *Canonical* if $\Theta = \mathrm{diag}(J, J, \ldots, J)$, or
- *Degenerate canonical* if $\Theta = \mathrm{diag}(0_{n' \times n'}, J, \ldots, J)$, where $0 < n' \leq n$.

Here, J denotes the real skew-symmetric 2×2 matrix

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix},$$

and the "diag" notation indicates a block diagonal matrix assembled from the given entries. To illustrate, the case of a system with one classical variable and two conjugate quantum variables is characterized by $\Theta = \mathrm{diag}(0, J)$, which is degenerate canonical.

Assume for the moment that Θ is canonical (hence n is even). Then the operators $x_1(0), \ldots, x_n(0)$ satisfying (7.3) and the underlying Hilbert space, say $\mathcal{H}_s$, can be realized in the standard way via the GNS construction and Stone's Theorem; for details see [97, 92]. Important in this construction are the so-called *Weyl operators* $\{W(x); x \in \mathbb{R}^n\}$ of bounded operator on $\mathcal{H}_s$. They satisfy the *Weyl relations* $W(x)W(y) = e^{-i\langle x, \Theta y\rangle} W(x+y)$ for all $x, y \in \mathbb{R}^n$. It is well-known that the Weyl operators generate the von Neumann algebra of all bounded operators on $\mathcal{H}_s$. Let us denote this algebra by $\mathscr{A}_s$ and assign to it a state $\mathbb{P}_s$ such that x_0 is Gaussian. We shall denote the density operator associated with $\mathbb{P}_s$ by ρ and say that x_0 *is Gaussian with state* ρ. If Θ is degenerate canonical, we first construct a so-called augmentation of (7.2), which is developed in Section 7.3.2, and perform the same construction as before on this augmentation to realize x_0 and $\mathcal{H}_s$.

The vector quantity w describes the input signals and is assumed to admit the decomposition

$$dw(t) = \beta_w(t)dt + d\tilde{w}(t) \tag{7.4}$$

where $\tilde{w}(t)$ is the noise part of $w(t)$ and $\beta_w(t)$ is a self adjoint, adapted process (see, e.g., [91, 92, 94] for a discussion of adapted processes).

The noise $\tilde{w}(t)$ is a vector of self-adjoint quantum noises with Ito table

$$d\tilde{w}(t)d\tilde{w}^T(t) = F_{\tilde{w}}dt, \tag{7.5}$$

where $F_{\tilde{w}}$ is a non-negative Hermitian matrix; e.g., see [92, 101]. This determines the following commutation relations for the noise components:

$$[d\tilde{w}(t), d\tilde{w}^T(t)] = d\tilde{w}(t)d\tilde{w}^T(t) - (d\tilde{w}(t)d\tilde{w}^T(t))^T = 2T_{\tilde{w}}dt, \tag{7.6}$$

where we use the notation $S_{\tilde{w}} = \frac{1}{2}(F_{\tilde{w}} + F_{\tilde{w}}^T)$, $T_{\tilde{w}} = \frac{1}{2}(F_{\tilde{w}} - F_{\tilde{w}}^T)$ so that $F_{\tilde{w}} = S_{\tilde{w}} + T_{\tilde{w}}$. For instance, $F_{\tilde{w}} = \mathrm{diag}(1, I + iJ)$ describes a noise vector with one classical component and a pair of conjugate quantum noises (here I is the 2×2 identity matrix). The noise processes can be represented as operators on an appropriate Fock space (a particular, yet important, type of Hilbert space); e.g., see [91, 92]. Let us denote this noise Fock space by $\mathcal{F}$ and the algebra of all bounded operators on $\mathcal{F}$ by $\mathscr{W}$. We assume that $\mathscr{W}$ is assigned a Gaussian/quasi-free state ϕ such that the noise commutation relations (7.6) hold. Indeed, later in this section $F_{\tilde{w}}$ is assumed a certain canonical form corresponding to ϕ being a *vacuum state*.

The process $\beta_w(t)$ serves to represent variables of other systems which may be passed to the system (7.2) via a connection. Therefore, we require that $\beta_w(0)$ is an operator on a Hilbert space H_a distinct from $\mathcal{H}_s$ and $\mathcal{F}$. We also assume $\beta_w(t)$ commutes with $x(t)$ for all $t \geq 0$ (two vectors x, y of operators are said to commute if $xy^T - (yx^T)^T = 0$); this will simplify matters for the present work. Moreover, since we had earlier specified that $\beta_w(t)$ should be an adapted process, we make note that $\beta_w(t)$ also commutes with $d\tilde{w}(t)$ for all $t \geq 0$. We denote the von Neumann algebra of all bounded operators on $\mathcal{H}_a$ by $\mathcal{A}_a$ and assume that it is assigned a state $\mathbb{P}_a$.

Overall, the system (7.2) is defined on the composite Hilbert space $\mathcal{H}_c = \mathcal{H}_s \otimes \mathcal{H}_a \otimes \mathcal{F}$ and all operators are *affiliated* to the von Neumann algebra $\mathscr{A}_c = \mathscr{A}_s \otimes \mathscr{A}_a \otimes \mathscr{W}$ for all $t \geq 0$. A self-adjoint operator X is said to be affiliated to a von Neumann algebra $\mathscr{A}$ if $(X + iI)^{-1} \in \mathscr{A}$. Affiliation is a useful notion for relating *unbounded* operators, such as the components of $x(t)$, to an algebra of *bounded* operators. Statistical attributes of the operators are determined by the composite state $\mathbb{P}_c = \mathbb{P}_s \otimes \mathbb{P}_a \otimes \phi$. Therefore, the associated quantum probability space for (7.2) is $(\mathscr{A}_c, \mathbb{P}_c)$. Note here that operators originally defined on $\mathcal{H}_s$, $\mathcal{H}_a$ or $\mathcal{F}$ are implicitly "lifted" to the composite space $\mathcal{H}_c$ by the standard operation of ampliation (i.e., tensoring with appropriate identity operators). For example, the ampliation of an operator $X : \mathcal{H}_s \to \mathcal{H}_s$ to $\mathcal{H}_c$ is simply $X \otimes I \otimes I$ where the middle I denotes the identity operator on $\mathcal{H}_a$ while the right most I denotes the identity operator on $\mathcal{F}$.

At this point, we stress once again that the most important fact to be noted of the model (7.2) is that its similarity in form to the state-space model of classical finite-dimensional stochastic linear systems. The *main difference* is that in the classical setting, the vectors $x(t)$, $y(t)$ and $w(t)$ consist of real- or complex-valued

functions of time which *commute* with one another for all time $t \geq 0$ and different times $s, t \geq 0$, while in the quantum context they consist of operators on some Hilbert space which *need not commute* with one another at any time $t \geq 0$ nor at any two time instances $s, t \geq 0$. But despite this difference, we shall show in the next chapter that in the context of quantum H^∞ control, to some degree it is possible to work with them in a similar way as we do with their classical counterparts.

To simplify the exposition, we now set up some conventions to put the system (7.2) into a standard form. First, note that there will be no change to the dynamics of $x(t)$ and $y(t)$ if we enlarge $w(t)$, by adding additional dummy noise components and enlarging $\mathcal{F}$ and $\mathscr{W}$ if necessary, and at the same time enlarging B by inserting suitable columns of zeros. Secondly, we may add dummy components to y by enlarging C and D by inserting additional dummy rows to each of these matrices. Our original output can be recovered by discarding or "disconnecting" the dummy components/entries. Therefore, we make the following assumptions on the system (7.2): (i) n_y is even, and (ii) $n_w \geq n_y$. We also make the assumption that $F_{\tilde{w}}$ is of the *canonical* form $F_{\tilde{w}} = I + i\mathrm{diag}(J, \ldots, J)$. Hence n_w has to be even. Note that if $F_{\tilde{w}}$ is not canonical but of the form $F_{\tilde{w}} = I + i\mathrm{diag}(0_{n' \times n'}, \mathrm{diag}(J, \ldots, J))$ with $n' \geq 1$, we may enlarge $w(t)$ (and hence also $\tilde{w}(t)$) and B as before such that the enlarged noise vector, say $\tilde{w}'$, can be taken to have an Ito matrix $F_{\tilde{w}'}$ which is canonical.

Equation (7.2) is a linear quantum stochastic differential equation. General quantum stochastic differential equations of this type are described in [91, 92], though the specific linear equations (7.2) may be treated directly, with solutions given explicitly by:

$$\begin{aligned} x(t) &= e^{At}x(0) + \int_0^t e^{A(t-s)}Bdw(s); \\ y(t) &= \int_0^t Cx(s)ds + Dw(t). \end{aligned} \tag{7.7}$$

Here the integral with respect to $dw(t)$ is taken to be a quantum stochastic integral. By construction, $x(t)$ depends only on the past noise $w(s)$, for $0 \leq s \leq t$; i.e., it is adapted, and a property of the Ito increments is that $dw(t)$ commutes with $x(t)$.

Equations (7.2) describe a general non-commutative linear stochastic system, which need not necessarily correspond to a physical system, such as an optical cavity, attenuator, or amplifier. This issue does not normally arise in physical

modelling, but we shall see in the next chapter that it is of considerable importance when we come to synthesizing *physically realizable* controllers. In particular, we will describe a quantum H^∞ control framework for quantum linear stochastic systems due to James and Petersen [32]. This framework returns a partial model for a quantum linear stochastic controller which may not represent any physically meaningful system. The main idea is to complete the partial model by suitably adding additional channels of quantum Wiener processes to the partial model such that the completed model does indeed represent some physically meaningful system. In the next section, we shall formalize what we mean by "physically meaningful" by introducing a precise notion of physical realizability for quantum linear stochastic systems represented by the QSDE (7.2).

7.3 Physical realizability of linear QSDEs

As mentioned at the beginning of this chapter, a basic requirement of any physical quantum system is that canonical commutation relations (CCR) between canonical operators of the system must be preserved for all time $t \geq 0$ (in the Heisenberg or interaction picture of quantum mechanics [96]). In our context, the canonical operators are elements of the vector $x(t)$ and the preservation of the CCR translates to the condition:

$$[x_i(t), x_j(t)] = 2i\Theta_{ij} \quad \forall i, j = 1, \ldots, n \text{ and } \forall t \geq 0,$$

or equivalently,

$$x(t)x(t)^T - (x(t)x(t)^T)^T = 2i\Theta \quad \forall t \geq 0.$$

The following theorem provides an algebraic characterization of precisely when the quantum system (7.2) preserves the commutation relations as time evolves.

Theorem 7.3.1 *Under the assumptions of Section 7.2 for the system (7.2),* -$[x_i(0), x_j(0)] = 2i\Theta_{ij}$ *implies* $[x_i(t), x_j(t)] = 2i\Theta_{ij}$ *for all* $t \geq 0$ *if and only if*

$$iA\Theta + i\Theta A^T + BT_{\tilde{w}}B^T = 0. \tag{7.8}$$

Proof. To preserve the commutation relations for all $i, j = 1, \ldots, n$ and all $t \geq 0$, we must have $d[x_i, x_j] = 0$ for all $i, j = 1, \ldots, n$ (for convenience, in this proof we shall drop the time index t). We now develop a general expression for $d[x_i, x_j]$. Indeed, let $e_k = [\ 0 \quad \ldots \quad 0 \quad 1 \quad 0 \quad \ldots \quad 0\ ]^T$, where the 1 is in the k-th

row. It is easy to see that for any $i, j \in \{1, \ldots, n\}$, $[x_i, x_j] = e_i^T x x^T e_j - e_j^T x x^T e_i$. Therefore, $d[x_i, x_j] = e_i^T d(xx^T) e_j - e_j^T d(xx^T) e_i$. Now, we expand $d(xx^T)$ using the quantum Ito rule (e.g., see [92]) as follows:

$$\begin{aligned} d(xx^T) &= (dx)x^T + xd(x^T) + dxd(x^T) \\ &= Axx^T dt + Bdwx^T + xx^T A^T dt + xd(w^T)B^T + Axx^T A^T dt^2 \\ &\quad + Axd(w^T)dtB^T + Bdwdtx^T A^T + Bdw(dw)^T B^T \\ &= Axx^T dt + Bdwx^T + xx^T A^T dt + xd(w^T)B^T + B(dw)(dw)^T B^T. \end{aligned}$$

Substituting $dw = \beta_w dt + d\tilde{w}$ into the above and noting that $\beta_w \beta_w^T dt^2$ and $\beta_w d\tilde{w}^T dt$ vanish to order dt gives

$$\begin{aligned} d(xx^T) &= Axx^T dt + B\beta_w x^T dt + Bd\tilde{w}x^T + xx^T A^T dt + x\beta_w^T B^T dt + xd\tilde{w}^T B^T + \\ &\quad Bd\tilde{w}d\tilde{w}^T B^T. \end{aligned}$$

We now write $A = [\ A_1^T \quad A_2^T \quad \ldots \quad A_n^T\ ]^T$ and $B = [\ B_1^T \quad B_2^T \quad \ldots \quad B_n^T\ ]^T$, where the vectors A_k and B_k denote the k-th row of matrices A and B, respectively. Then we have

$$\begin{aligned} e_i^T d(xx^T) e_j &= e_i^T Axx^T e_j dt + e_i^T B\beta_w x^T e_j dt + e_i^T Bd\tilde{w}x^T e_j + e_i^T xx^T A^T e_j dt \\ &\quad + e_i x\beta_w^T B^T e_j dt + e_i xd\tilde{w}^T B^T e_j + e_i^T Bd\tilde{w}(d\tilde{w})^T B^T e_j \\ &= A_i x x_j dt + B_i \beta_w x_j dt + B_i d\tilde{w} x_j + x_i A_j x dt + x_i B_j \beta_w dt + x_i B_j d\tilde{w} + \\ &\quad (B_i d\tilde{w})(B_j d\tilde{w}). \end{aligned} \tag{7.9}$$

Also we have

$$\begin{aligned} e_j^T d(xx^T) e_i &= A_j x x_i dt + B_j \beta_w x_i dt + B_j d\tilde{w} x_i + x_j A_i x dt + \\ &\quad x_j B_i \beta_w dt + x_j B_i d\tilde{w} + (B_j d\tilde{w})(B_i d\tilde{w}). \end{aligned} \tag{7.10}$$

Subtracting (7.10) from (7.9) gives us

$$\begin{aligned} & e_i^T d(xx^T) e_j - e_j^T d(xx^T) e_i \\ &= ((A_i x)x_j - x_j(A_i x))dt + ((B_i\beta_w)x_j - x_j(B_i\beta_w))dt \\ &\quad + (B_i d\tilde{w})x_j - x_j(B_i d\tilde{w}) + (x_i(A_j x) - (A_j x)x_i)dt \\ &\quad + (x_i(B_j\beta_w) - (B_j\beta_w)x_i)dt + (x_i(B_j d\tilde{w}) - (B_j d\tilde{w})x_i) \\ &\quad + ((B_i d\tilde{w})(B_j d\tilde{w}) - (B_j d\tilde{w})(B_i d\tilde{w})) \\ &= ((A_i x)x_j - x_j(A_i x))dt + ((B_i\beta_w)x_j - x_j(B_i\beta_w))dt \\ &\quad + (x_i(A_j x) - (A_j x)x_i)dt + (x_i(B_j\beta_w) - (B_j\beta_w)x_i)dt \\ &\quad + ((B_i d\tilde{w})(B_j d\tilde{w}) - (B_j d\tilde{w})(B_i d\tilde{w})). \end{aligned} \tag{7.11}$$

Here we are using the fact that elements of $d\tilde{w}$ commute with those of x and β_w due to the adaptedness of x and β_w. Hence,

$$\begin{aligned}
& e_i^T d(xx^T)e_j - e_j^T d(xx^T)e_i \\
&= [A_i x, x_j]dt - [x_j, B_i\beta_w]dt + [x_i, A_j x]dt + [x_i, B_j\beta_w]dt + \\
&\quad [B_i d\tilde{w}, B_j d\tilde{w}] \\
&= \sum_{k=1}^n A_{ik}[x_k, x_j]dt - \sum_{k=1}^n B_{ik}[x_j, \beta_{wk}]dt + \sum_{k=1}^n A_{jk}[x_i, x_k]dt \\
&\quad + \sum_{k=1}^n B_{jk}[x_i, \beta_{wk}]dt + \sum_{k=1}^n \sum_{l=1}^n B_{ik}B_{jl}[d\tilde{w}_k, d\tilde{w}_l] \\
&= \left(2i\sum_{k=1}^n A_{ik}\Theta_{kj} + 2i\sum_{k=1}^n A_{jk}\Theta_{ik} - \sum_{k=1}^n B_{ik}C_{jk}^{x\beta_w} + \sum_{k=1}^n B_{jk}C_{ik}^{x\beta_w} \right. \\
&\quad \left. + \sum_{k=1}^n \sum_{l=1}^n B_{ik}B_{jl}(F_{\tilde{w},kl} - F_{\tilde{w},lk}) \right) dt, \qquad (7.12)
\end{aligned}$$

where $C_{ij}^{x\beta_w} = [x_i, \beta_{wj}]$. Since $C^{x\beta_w}=[C_{ij}^{x\beta_w}]_{i=1,\ldots,n,j=1\ldots,n_{\beta_w}} = 0$ (by assumption) and $F_{\tilde{w}} - F_{\tilde{w}}^T = 2T_{\tilde{w}}$, equation (7.12) takes the form

$$d(xx^T - (xx^T)^T) = 2(iA\Theta + i\Theta A^T + BT_{\tilde{w}}B^T)dt \qquad (7.13)$$

from which the result follows. □

Thus we see that preservation of the CCR amounts to an algebraic constraint (7.8) that must be satisfied by the system matrices A and B of (7.2). However, as we shall see shortly, for (7.2) to be physically realizable there is actually another constraint required in relation to the output signal $y(t)$. We shall now proceed further by introducing the notion of an *open quantum harmonic oscillator*, which acts as the basic "dynamical unit" (as opposed to static quantum optical components units/devices such as beamsplitters and phase shifters) of a physically realizable quantum system.

7.3.1 Open quantum harmonic oscillator

In order to formally present a definition of an open quantum harmonic oscillator we will require the following notation. For a square matrix T, $\mathrm{diag}_m(T)$ denotes the block diagonal matrix $\mathrm{diag}(T, \ldots, T)$ where T appears m times as a diagonal block. The symbol P_m denotes a $2m \times 2m$ *permutation matrix* defined so that if we consider a column vector $a = [\ a_1 \ \ a_2 \ \ \ldots \ \ a_{2m}\ ]^T$, then

$P_m a = [\ a_1 \quad a_3 \quad \dots \quad a_{2m-1} \quad a_2 \quad a_4 \quad \dots \quad a_{2m}\ ]^T$. Recall that an $m \times m$ permutation matrix is a full-rank real matrix whose columns (or, equivalently, rows) consist of standard basis vectors for $\mathbb{R}^m$; i.e., vectors in $\mathbb{R}^m$ whose elements are all 0 except for one element which has the value 1. A permutation matrix P has the unitary property $PP^T = P^T P = I$. Note that $P_m^T[\ a_1 \quad a_2 \quad \dots \quad a_{2m}\ ]^T = [\ a_1 \quad a_{m+1} \quad a_2 \quad a_{m+2} \quad \dots \quad a_m \quad a_{2m}\ ]^T$.

Let us also further introduce the notation $N_w = \frac{n_w}{2}$ and $N_y = \frac{n_y}{2}$,

$$M = \frac{1}{2}\begin{bmatrix} 1 & i \\ 1 & -i \end{bmatrix},$$

and $\Gamma = P_{N_w}\mathrm{diag}_{N_w}(M)$. Moreover, let * denote the adjoint of a Hilbert space operator (by this we mean that the operator is a map from one Hilbert space to another), and let $X^\#$ denote the operation of taking the adjoint of each element of X, where X is a matrix/array of Hilbert space operators. Also, let $X^\dagger = (X^\#)^T$.

Then we have the following definition of a quantum harmonic oscillator by slightly generalizing a linear model given in [102, Section 4]:

Definition 7.3.2 *Set $\beta_w(t) = 0\ \forall t \geq 0$. Then the system (7.2) is said to be an* open quantum harmonic oscillator *if Θ is canonical and there exist a quadratic Hamiltonian $H = x(0)^T R x(0)$, with a real and symmetric Hamiltonian matrix R of dimension $n \times n$, and a coupling operator $L = \Lambda x(0)$, with complex-valued coupling matrix Λ of dimension $n_w \times n$, such that:*

$$x(t) = U(t)^* x(0) U(t), \quad y_l(t) = U(t)^* w_l(t) U(t), \quad l = 1, \dots, n_y,$$

where $\{U(t); t \geq 0\}$ is an adapted process of unitary operators satisfying the following QSDE [102, Section 2.5] :

$$dU(t) \ = \ (-iH dt - \frac{1}{2}L^\dagger L dt + [\ -L^\dagger\ L^T\ ]\Gamma dw(t))U(t),\ U(0) = I.$$

In this case the matrices A, B, C, D are given by:

$$A = 2\Theta(R + \Im(\Lambda^\dagger \Lambda)); \tag{7.14}$$

$$B = 2i\Theta[\ -\Lambda^\dagger\ \Lambda^T\ ]\Gamma; \tag{7.15}$$

$$C = P_{N_y}^T \begin{bmatrix} \Sigma_{N_y} & 0_{N_y \times N_w} \\ 0_{N_y \times N_w} & \Sigma_{N_y} \end{bmatrix} \begin{bmatrix} \Lambda + \Lambda^\# \\ -i\Lambda + i\Lambda^\# \end{bmatrix}; \tag{7.16}$$

$$D = P_{N_y}^T \begin{bmatrix} \Sigma_{N_y} & 0_{N_y \times N_w} \\ 0_{N_y \times N_w} & \Sigma_{N_y} \end{bmatrix} P_{N_w} = [\ I_{n_y \times n_y} \quad 0_{n_y \times (n_w - n_y)}\ ], \tag{7.17}$$

where $\Sigma_{N_y} = [\ I_{N_y \times N_y}\ 0_{N_y \times (N_w - N_y)}\ ]$.

As its name suggests, the open quantum harmonic oscillator is simply an $n-$dimensional quantum harmonic oscillator with Hamiltonian H which is *open* in the sense that it interacts with its environment (for a discussion of open quantum systems, see [103], [99, Chapter 3]), which in this case are independent quantum Wiener noise channels, and the interaction with the environment is linear via the coupling operator L. Note that in quantum optics, a quantum Wiener noise channel is an idealized model a of a free travelling quantized electromagnetic field, which is precisely how these channels can be physically realized in the laboratory. For a discussion see, for example, [93, 86].

Remark 7.3.3 *In the definition, we have set $\beta_w \equiv 0$ since an open quantum harmonic oscillator is a stand alone open system in its ambient heat bath, which in this case are the quantum noise channels. Recall that β_w serves to represent observables originating from another physical system via an interconnection.*

Another important point to be noted in the definition of an open quantum harmonic oscillator is that the output $y(t)$ has a specific form, that is, it is the time evolved version of the noise channels $w(t)$ after its interaction with the oscillator (via the unitary evolution U_t). This can be considered a natural restriction because observables of a physical system of interest (such as an optical cavity or an atom) cannot be observed directly, but only indirectly such as by shining a laser or light source on the system and observing the light which is reflected as a result of the interaction of the incident light with the system. Here, $w(t)$ plays the role of the incident light, while $y(t)$ is the reflected light. A intuitive visualization of the concept an open quantum harmonic oscillator is given in Figure 7.1.

7.3.2 Augmentation of a linear QSDE

If Θ is degenerate canonical then we may perform an augmentation in which Θ is embedded into a larger skew symmetric matrix $\tilde{\Theta}$ which is canonical *up to permutation* (this means $\tilde{\Theta}$ becomes canonical after permutation of appropriate rows and columns). To do this, let $\theta = [\Theta_{ij}]_{i,j=n'+1,\ldots,n} = \mathrm{diag}_{\frac{n-n'}{2}}(J)$ if $n' < n$. Here $\mathrm{diag}_m(J)$ denotes a $m \times m$ block diagonal matrix with m matrices J on the diagonal. Define:

$$\tilde{\Theta} = \begin{bmatrix} 0_{n'\times n'} & 0_{n'\times(n-n')} & I_{n'\times n'} \\ 0_{(n-n')\times n'} & \theta & 0_{(n-n')\times n'} \\ -I_{n'\times n'} & 0_{n'\times(n-n')} & 0_{n'\times n'} \end{bmatrix},$$

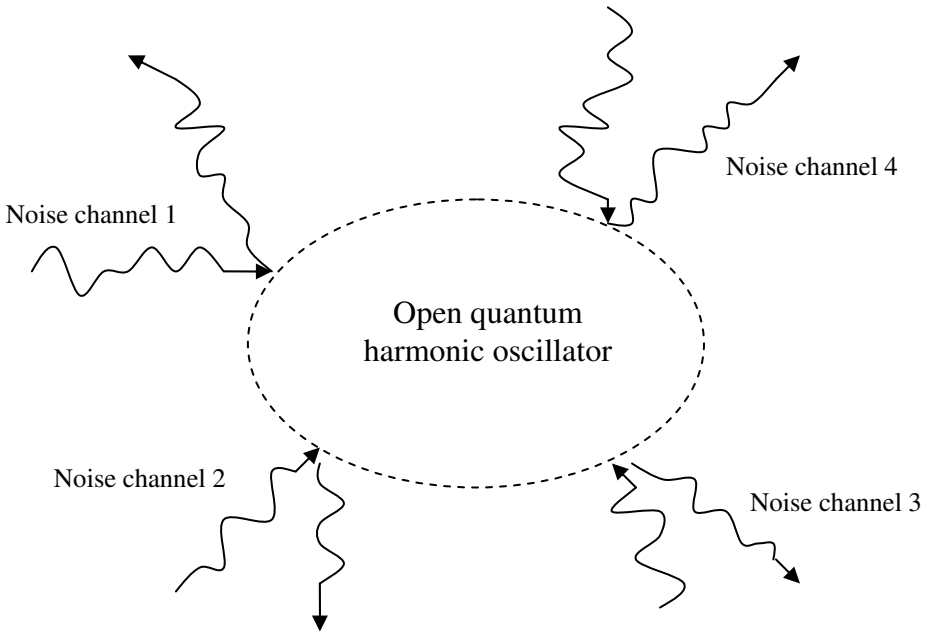

Figure 7.1: Visualization of an open quantum harmonic oscillator

where the middle block of rows is dropped whenever $n = n'$. Then by definition $\tilde{\Theta}$ is canonical up to permutation and contains Θ as a sub-matrix by removing appropriate rows and columns of $\tilde{\Theta}$. Let $\tilde{n} = n + n'$, the dimension of the rows and columns of $\tilde{\Theta}$. By enlarging if necessary the quantum probability space, define the vector $\tilde{x}(t) = [x_1(t)\, x_2(t)\, \ldots\, x_n(t)\, z_1(t)\, z_2(t)\, \ldots\, z_{n'}(t)]^T$ of variables. We now define the following linear QSDE

$$\begin{aligned} d\tilde{x}(t) &= \begin{bmatrix} A & 0_{n\times n_c} \\ A' & A'' \end{bmatrix} \tilde{x}(t)dt + \begin{bmatrix} B \\ B' \end{bmatrix} dw(t), \\ \tilde{y}(t) &= \begin{bmatrix} C & C' \end{bmatrix} \tilde{x}(t)dt + Ddw(t) \end{aligned} \tag{7.18}$$

where A', A'', B' and C' are, respectively, some real $n' \times n$, $n' \times n'$, $n' \times n_w$ and $n_y \times n'$ matrices, and the initial variables $\tilde{x}(0)$ satisfy the commutation relations $\tilde{x}_0\tilde{x}(0)^T - (\tilde{x}(0)\tilde{x}(0)^T)^T = 2i\tilde{\Theta}$. We shall refer to the system (7.18) as an *augmentation* of (7.2).

Remark 7.3.4 *In the proof of Theorem 7.3.6 it is shown that the augmentation can be chosen to preserve commutation relations whenever the original system does.*

An augmentation is useful for handling vectors $x(t)$ which may contain one or more classical components, i.e., when Θ is degenerate canonical. The central idea

here is that in an augmentation, any classical component of $x(t)$, say $x_k(t)$, is considered to be "one-half" of a pair of canonically conjugate operators $\{x_k(t), z_k(t)\}$ satisfying $[x_k(t), z_k(t)] = 2i$, where $z_k(t)$ also satisfies $[z_k(t), x_l(t)] = 0 \ \forall l \neq k$. Here $z_k(t)$ acts merely as a "dummy variable" which becomes a component of the augmented vector $\tilde{x}(t)$, in the sense that $z(t)$ has no effect whatsoever on the dynamics of $x(t)$, as can be seen from the QSDE (7.18). Thus, here augmentation presents a convenient way of treating classical components within the formalisms of quantum mechanics, without having to develop new concepts or theories for handling them. The fact that a quantum linear stochastic system which preserve the CCR is guaranteed to have an augmentation which again preserve the CCR, as stated in Remark 7.3.4, shows that classical components can be "embedded" in a fully quantum mechanical augmentation in a consistent way.

7.3.3 Formal definition of physical realizability

With open quantum harmonic oscillators and augmentations having been defined, we are now ready to introduce a formal definition of physical realizability of the QSDE (7.2). A discussion regarding the definition follows after Theorem 7.3.6 in which necessary and sufficient conditions for physical realizability are given.

Definition 7.3.5 *The system (7.2) is said to be physically realizable if one of the following holds:*

1. Θ *is canonical and (7.2) represents the dynamics of an open quantum harmonic oscillator.*

2. Θ *is degenerate canonical and there exists an augmentation (7.18) which, after a suitable relabelling of the components $\tilde{x}_1(t), \ldots, \tilde{x}_{\tilde{n}}(t)$ of $\tilde{x}(t)$, represents the dynamics of an open quantum harmonic oscillator.*

The following theorem provides explicit necessary and sufficient conditions for physical realizability given in terms of the system matrices A, B, C, D. Hence, whether a given system (7.2) is physically realizability in the sense of Definition 7.3.5 can be determined in definite manner (i.e., it is either 'yes' or 'no'), which is a useful practical result.

Theorem 7.3.6 *The system (7.2) is physically realizable if and only if:*

$$iA\Theta + i\Theta A^T + BT_wB^T = 0, \tag{7.19}$$

$$B\begin{bmatrix} I_{n_y\times n_y} \\ 0_{(n_w-n_y)\times n_y}\end{bmatrix} = \Theta C^T P_{N_y}^T \begin{bmatrix} 0_{N_y\times N_y} & I_{N_y\times N_y} \\ -I_{N_y\times N_y} & 0_{N_y\times N_y}\end{bmatrix} P_{N_y} = \Theta C^T \mathrm{diag}_{N_y}(J), \tag{7.20}$$

and D satisfies (7.17). Moreover for canonical Θ, the Hamiltonian and coupling matrices have explicit expressions as follows. The Hamiltonian matrix R is uniquely given by $R = \frac{1}{4}(-\Theta A + A^T\Theta)$, and the coupling matrix Λ is given uniquely by

$$\Lambda = -\frac{1}{2}i\begin{bmatrix} 0_{N_w\times N_w} & I_{N_w\times N_w}\end{bmatrix}(\Gamma^{-1})^T B^T\Theta. \tag{7.21}$$

In the case that Θ is degenerate canonical, a physically realizable augmentation of the system can be constructed to determine the associated Hamiltonian and coupling operators using the above explicit formulas.

Remark 7.3.7 *Note that the Hamiltonian and coupling operators are determined by (7.19), while conditions (7.17) and (7.20) relate to the required form of the output equation.*

Proof. (of Theorem 7.3.6) Let us first consider the case where Θ is canonical. If the system is realizable then (7.14)-(7.17) holds. Since $U(t)$ is unitary for each $t \geq 0$, we have that $d\left(x(t)x(t)^T - (x(t)x(t)^T)^T\right) = 0$; i.e., the canonical commutation relations are preserved. By Theorem 7.3.1 this is equivalent to (7.19). Let $M_1, M_2, \ldots, M_{N_y}$ be column vectors such that $[M_1\ M_2\ \ldots\ M_{N_y}] = \Lambda^T[\ I_{N_y\times N_y}\ \ 0\ ]^T$. Then using (7.15) and (7.16) we obtain the following after some algebraic manipulations:

$$\begin{aligned} B[\ I_{n_y\times n_y}\ \ 0_{n_y\times(n_w-n_y)}\ ]^T &= 2i\Theta[\ -\Lambda^\dagger\ \ \Lambda^T\ ]\Gamma[\ I_{n_y\times N_y}\ \ 0_{n_y\times(n_w-N_y)}\ ]^T \\ &= 2\Theta[-\Im(M_1)\,\Re(M_1)\ \ldots\ -\Im(M_{N_y})\,\Re(M_{N_y})] \\ &= \Theta\left(P_{N_y}^T\begin{bmatrix} 0_{N_y\times N_y} & -I_{N_y\times N_y} \\ I_{N_y\times n_y} & 0_{N_y\times n_y}\end{bmatrix}P_{N_y}C\right)^T \\ &= \Theta C^T P_{N_y}^T\begin{bmatrix} 0 & I_{N_y\times N_y} \\ -I_{N_y\times N_y} & 0\end{bmatrix}P_{N_y} \\ &= \Theta C^T\mathrm{diag}_{N_y}(J). \end{aligned}$$

Therefore, we conclude that (7.20), (7.19) and (7.17) are necessary for realizability.

Conversely, now suppose that (7.20), (7.19) and (7.17) hold. We will argue that these conditions are sufficient for realizability by showing that they imply the existence a symmetric matrix R and a coupling matrix Λ such that (7.14)-(7.16) are satisfied. First we note that after some simple algebraic manipulation $-i\Theta^{-1}B\Gamma^{-1} = i\Theta B\Gamma^{-1} = [-Z^{\#}\ Z]$, for some complex matrix Z. Hence $B = i\Theta[-Z^{\#}\ Z]\Gamma$. Substituting the last expression into (7.19) and after further manipulations we get:

$$iA\Theta + i\Theta A^T - \frac{1}{2}\Theta(Z^{\#}Z^T - ZZ^{\dagger})\Theta = 0.$$

Writing $Z^{\#}Z^T - ZZ^{\dagger} = 2i\Im(Z^{\#}Z^T)$, we may rewrite the last expression as follows:

$$\begin{aligned} iA\Theta + i\Theta A^T - \frac{1}{2}\Theta(Z^{\#}Z^T - ZZ^{\dagger})\Theta &= iA\Theta + i\Theta A^T - i\Theta\Im(Z^{\#}Z^T)\Theta \\ &= i\Theta(\Theta^{-1}A + A^T\Theta^{-1} - \Im(Z^{\#}Z^T))\Theta \\ &= i\Theta(\Theta^{-1}A - (\Theta^{-1}A)^T - \Im(Z^{\#}Z^T))\Theta \\ &= 0, \end{aligned}$$

implying that $\Theta^{-1}A - (\Theta^{-1}A)^T - \Im(Z^{\#}Z^T) = 0$. Since $\Theta^{-1}A$ is real, we have the decomposition $\Theta^{-1}A = -\Theta A = V + W$ for a unique pair of real symmetric matrix V and real skew symmetric matrix W and obtain the condition $2W - \Im(Z^{\#}Z^T) = 0$. Hence, $W = \frac{1}{2}\Im(Z^{\#}Z^T)$. Setting $R = \frac{1}{2}V$ and $\Lambda = 2Z^T$, we get $A = 2\Theta(R + \Im(\Lambda^{\dagger}\Lambda))$ and $B = 2i\Theta[-\Lambda^{\dagger}\ \Lambda^T]$ as desired, and also prove the second statement of the theorem. After substituting the expression, just obtained for B (in terms of Λ, Θ, and Γ) into (7.20) and more algebraic manipulations we then get (7.16). Since the expression for D has been hypothesized as (7.17), we conclude that (7.20), (7.19) along with (7.17) gives matrices A, B, C, D which are the coefficients of a realizable system.

Now, we consider the case where Θ is degenerate canonical, i.e.,

$$\Theta = \mathrm{diag}(0_{n'\times n'}, \mathrm{diag}_{\frac{n-n'}{2}}(J)).$$

Let us write

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \quad B = \begin{bmatrix} B_1 & B_2 \end{bmatrix} \quad C = \begin{bmatrix} C_1 & C_2 \end{bmatrix}$$

with $A_{11} \in \mathbb{R}^{n' \times n'}$, $A_{12} \in \mathbb{R}^{n' \times (n-n')}$, $A_{21} \in \mathbb{R}^{(n-n') \times n'}$, $A_{22} \in \mathbb{R}^{(n-n') \times (n-n')}$, $B_1 \in \mathbb{R}^{n \times n_y}$, $B_2 \in \mathbb{R}^{n \times (n_w - n_y)}$, $C_1 \in \mathbb{R}^{n_y \times n'}$ and $C_2 \in \mathbb{R}^{n_y \times (n-n')}$. Consider the following augmentation:

$$\begin{aligned} d\tilde{x}(t) &= \begin{bmatrix} A_{11} & A_{12} & 0_{n' \times n'} \\ A_{21} & A_{22} & 0_{(n-n') \times n'} \\ A_1' & A_2' & A'' \end{bmatrix} \tilde{x}(t)dt + \begin{bmatrix} B_1 & B_2 \\ B_1' & 0 \end{bmatrix} dw(t) \\ d\tilde{y}(t) &= \begin{bmatrix} C & 0_{n_y \times n'} \end{bmatrix} \tilde{x}(t)dt + Ddw(t) \end{aligned}$$

where $B_1' = -C_1^T P_{N_y}^T \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P_{N_y}$, and A_1', A_2' and A'' satisfy the following:

$$\begin{aligned} A_1' - (A_1')^T &= i \begin{bmatrix} B_1' & 0 \end{bmatrix} T_w \begin{bmatrix} (B_1')^T \\ 0 \end{bmatrix} \\ \begin{bmatrix} A'' & -A_2' \mathrm{diag}_{\frac{n-n'}{2}}(J) \end{bmatrix} &= -\begin{bmatrix} A_{11}^T & A_{21}^T \end{bmatrix} - i \begin{bmatrix} B_1' & 0 \end{bmatrix} T_w B^T. \end{aligned}$$

It follows by inspection that such matrices A_1', A_2' and A'' exist. Let $A' = [A_1' \; A_2']$ and define

$$\tilde{A} = \begin{bmatrix} A & 0_{n \times n'} \\ A' & A'' \end{bmatrix} \quad \tilde{B} = \begin{bmatrix} B_1 & B_2 \\ B_1' & 0 \end{bmatrix} \quad \tilde{C} = \begin{bmatrix} C & 0_{n \times n'} \end{bmatrix}.$$

If (7.19) holds then it can be verified, by direct substitution, that the matrices $\tilde{A}$ and $\tilde{B}$ satisfy:

$$i\tilde{A}\tilde{\Theta} + i\tilde{\Theta}\tilde{A}^T + \tilde{B}T_w\tilde{B}^T = 0. \tag{7.22}$$

Recalling that $\tilde{\Theta}$ is only canonical up to permutation, we now need to transform it into canonical form. To do this, introduce the variable $z = P\tilde{x}$ where P is a permutation matrix such that $P\tilde{\Theta}P^T = \mathrm{diag}_{\frac{\tilde{n}}{2}}(J)$. Then the components of z are a relabelling of the components of $\tilde{x}$. This gives us the following dynamics for z:

$$\begin{aligned} dz(t) &= P\tilde{A}P^T z(t) + P\tilde{B}dw(t) \\ dy(t) &= \tilde{C}P^T z(t)dt + Ddw(t). \end{aligned}$$

Denoting $\hat{A} = P\tilde{A}P^T$, $\hat{B} = P\tilde{B}$, $\hat{C} = \tilde{C}P^T$, and $\hat{\Theta} = \mathrm{diag}_{\frac{\tilde{n}}{2}}(J)$ we see that (7.22) implies that:

$$i\hat{A}\hat{\Theta} + i\hat{\Theta}\hat{A}^T + \hat{B}T_w\hat{B}^T = 0. \tag{7.23}$$

Continuing further using (7.20), we have the following:

$$\begin{aligned}
\hat{B}\begin{bmatrix} I_{n_y\times n_y} \\ 0_{(n_w-n_y)\times n_y}\end{bmatrix} &= P\begin{bmatrix} B & \\ B_1' & 0\end{bmatrix}\begin{bmatrix} I_{n_y\times n_y} \\ 0_{(n_w-n_y)\times n_y}\end{bmatrix} \\
&= P\begin{bmatrix} \Theta C^T \\ -C_1^T\end{bmatrix} P_{N_y}^T\begin{bmatrix} 0 & I \\ -I & 0\end{bmatrix} P_{N_y} \\
&= P\tilde{\Theta}\begin{bmatrix} C^T \\ 0\end{bmatrix} P_{N_y}^T\begin{bmatrix} 0 & I \\ -I & 0\end{bmatrix} P_{N_y} \\
&= (P\tilde{\Theta}P^T)P\begin{bmatrix} C^T \\ 0\end{bmatrix} P_{N_y}^T\begin{bmatrix} 0 & I \\ -I & 0\end{bmatrix} P_{N_y} \\
&= \hat{\Theta}\,\hat{C}^T P_{N_y}^T\begin{bmatrix} 0 & I \\ -I & 0\end{bmatrix} P_{N_y} = \hat{\Theta}\hat{C}^T\mathrm{diag}_{N_y}(J). \quad (7.24)
\end{aligned}$$

If D is given by (7.17) then (7.23) and (7.24) implies, as we have already shown for the case of canonical Θ, the system defined by the matrices $(\hat{A}, \hat{B}, \hat{C}, D)$ is realizable in the sense of Point 1 of the theorem. Hence, the original system defined by the matrices (A, B, C, D) is then realizable in the sense of Point 2 of the theorem.

Finally, suppose conversely that (7.2) is realizable and let $(\tilde{A}, \tilde{B}, \tilde{C}, D)$ be a suitable augmentation. Then $(P\tilde{A}P^T, P\tilde{B}, \tilde{C}P^T, D)$ is a quantum harmonic oscillator, with P as defined before. Hence, $P\tilde{A}P^T$, $P\tilde{B}$, $\tilde{C}P^T$, and D are given by the right hand sides of (7.14)-(7.17) for a canonical Θ and some R and Λ. It follows that $\tilde{A}$, $\tilde{B}$, $\tilde{C}$ and $\tilde{D}$ are given by the same set of equations by replacing $\tilde{\Theta}$, R and Λ by $\tilde{\Theta} = P^T\Theta P$, $\tilde{R} = P^T R P$ and $\tilde{\Lambda} = \Lambda P$, respectively. We then have, from the same line of arguments given for the case of canonical Θ, that:

$$\tilde{B}\begin{bmatrix} I_{n_y\times n_y} \\ 0_{(n_w-n_y)\times n_y}\end{bmatrix} = \tilde{\Theta}\tilde{C}^T P_{N_y}^T\begin{bmatrix} 0 & I_{N_y\times N_y} \\ -I_{N_y\times N_y} & 0\end{bmatrix} P_{N_y} = \tilde{\Theta}\tilde{C}^T\mathrm{diag}_{N_y}(J), \quad (7.25)$$

(7.22) holds, and D satisfies (7.17). Reading off the first n rows of both sides of (7.25) then gives us (7.20), while reading of the first n rows and columns of both sides of (7.22) gives us (7.19), as required. This completes the proof. □

The conditions of Theorem 7.3.6 are precisely what we would intuitively expect. As stated in Theorem 7.3.1, (7.19) is the condition for preservation of the CCR, as required in a physical system. On the other hand, conditions (7.20) and (7.17) arise due to the restriction on the form of the output of an open quantum harmonic oscillator, as discussed at the end of Section 7.3.1.

We conclude this section with the following remark.

Remark 7.3.8 *It should be possible, and can be convenient, to consider the problem of physical realizability more broadly than discussed here by including additional static components, such as beam splitters and phase shifters that commonly occur in quantum optics (see [98, 57]). In a more general situation, one could consider output equations of the form:*

$$dy(t) = K_s(C_d x(t)dt + D_d dw(t)), \tag{7.26}$$

where K_s, C_d, D_d are real matrices satisfying $K_s C_d = C$ and $K_s D_d = D$. Here the matrix K_s represents the action of static devices connected to the output $\tilde{y}(t)$ of a physically realizable system (in the sense of Definition 7.3.5) defined by:

$$\begin{aligned} dx(t) &= Ax(t)dt + Bdw(t) \\ d\tilde{y}(t) &= C_d x(t)dt + D_d dw(t), \end{aligned}$$

where the quadruplet $\{A, B, C_d, D_d\}$ satisfy the conditions of Theorem 7.3.6. Note that $y(t) = K_s\tilde{y}(t)$. Therefore, in order that $y(t)$ has the correct Ito table, K_s should satisfy the constraint $K_s F_{\tilde{y}} K_s^T = I + \mathrm{diag}_{n_y}(J)$. However, detailed development of an efficient realization methodology, combining static and dynamical quantum units, for systems of the form (7.2) is beyond the scope of the thesis.

7.4 Concluding remarks

In this chapter, we have developed a notion of physical realizability of quantum linear stochastic systems which are relevant in quantum optics, and give explicit characterizations for physical realizability. In particular, we derive a necessary and sufficient condition for preservation of the canonical commutation relations for such systems, a prerequisite of any physical system.

Our results indicate that up to some degree one can work with quantum linear stochastic systems just as one would with classical linear stochastic systems. The key difference is that in the quantum case one has to take care of additional algebraic constraints imposed on the system matrices by Theorems 7.3.1 and 7.3.6. These new constraints are not present in the classical case.

The developments here support the point of view that it may be plausible to extend various controller synthesis methods that have been extensively developed over the years for classical linear stochastic systems to their quantum counterpart.

In the next chapter, we shall see that this is indeed the case for the well-known H^∞ synthesis methodology.

Chapter 8

Synthesis and Physical Realizability of H^∞ Quantum Linear Controllers

8.1 Introduction

Consider a given partial model of a quantum linear stochastic system, in which the system matrices A, C are completely specified, but the matrix B is only partially specified and the matrix D unspecified. This situation arises in the context of a H^∞ synthesis framework initiated by James and Petersen [32] for quantum linear stochastic systems. As mentioned briefly in the last chapter, it is then natural to ask whether one can complete the specification of B and determine D such that the resulting completely specified model is physically realizable in the sense of Chapter 7. James and Petersen have shown in some specific examples that it is indeed possible to complete the model, but the general case was an open question.

The purpose of this chapter is to address the general question of physical realizability of a partially specified quantum linear stochastic system. The main result is that given a partial model of a controller there always exists a physically realizable completion (Theorem 8.3.5). The results are constructive in the sense that we derive explicit formulas for B and D and the associated Hamiltonian and coupling matrices for the completion (Lemma 8.3.6). More importantly, however, is that our results show there is complete freedom to specify the commutation matrix Θ. This implies that in general one may in fact *a priori* specify the type

of controller to be realized. This means the controller can be chosen to be fully quantum, purely classical, or even a hybrid mixed classical-quantum controller. From an H^∞ perspective, the type of controller is inconsequential since the H^∞ performance is not affected by the particular choice of realization.

As with the previous chapter, this chapter is also based on the paper [100]. We begin with a discussion of the H^∞ synthesis framework of [32]. This will clearly show how the issue of physical realization of partially specified models arises in the controller synthesis for quantum linear stochastic systems.

8.2 Dissipation properties

In order to develop an H^∞ methodology for quantum linear systems, we first describe various dissipation properties frequently used in control engineering, suitably adapted to the quantum context. These properties concern the influence of disturbance inputs on energy transfers and stability. In particular, we give a quantum version of the Strict Bounded Real Lemma (Corollary 8.2.5) which will be employed in Section 8.3 for quantum H^∞ controller synthesis. In this section, we consider the following quantum system of the form (7.2):

$$\begin{aligned} dx(t) &= Ax(t)dt + \begin{bmatrix} B & G \end{bmatrix} \begin{bmatrix} dw(t)^T & dv(t)^T \end{bmatrix}^T; \\ dz(t) &= Cx(t)dt + \begin{bmatrix} D & H \end{bmatrix} \begin{bmatrix} dw(t)^T & dv(t)^T \end{bmatrix}^T \end{aligned} \tag{8.1}$$

In this quantum system, the input channel has two components, $dw = \beta_w dt + d\tilde{w}$ which represents the disturbance input, and dv which represents any additional noise input.

Definition 8.2.1 *Given an operator valued quadratic form*

$$r(x, \beta_w) = [x^T \beta_w^T] R \begin{bmatrix} x \\ \beta_w \end{bmatrix}$$

where

$$R = \begin{bmatrix} R_{11} & R_{12} \\ R_{12}^T & R_{22} \end{bmatrix}$$

is a given real symmetric matrix, we say the system (8.1) is dissipative with supply rate $r(x, \beta_w)$ *if there exists a positive operator valued quadratic form*

$V(x) = x^T X x$ (where X is a real positive definite symmetric matrix) and a constant $\lambda > 0$ such that

$$\langle V(x(t))\rangle + \int_0^t \langle r(x(s), \beta_w(s))\rangle ds \leq \langle V(x(0))\rangle + \lambda t \quad \forall t > 0, \tag{8.2}$$

for all Gaussian states ρ. Here we use the shorthand notation $\langle\cdot\rangle \equiv \mathbb{P}(\cdot)$ for expectation.

We say that the system (8.1) is strictly dissipative *if there exists a constant $\epsilon > 0$ such that inequality (8.2) holds with the matrix R replaced by the matrix $R + \epsilon I$.*

The term $\langle V(x(t))\rangle$ serves as the generalization to quantum stochastic systems (8.1) of the notion of the abstract internal energy for the system at time t. On the other hand, the term $\langle r(x(t), \beta_w(t))\rangle$ is a quantum generalization of the notion of abstract power flow into/out of the system at time t. Both of these are notions which are widely used in the stability analysis of linear and non-linear deterministic systems [104, 105]. The dissipation inequality (8.2) is a generalization of the corresponding inequality that was introduced for classical linear stochastic systems by Dupuis, James and Petersen [106]. Note that the term λt on the right hand side of (8.2), which accounts for the variance of Wiener process disturbances, pertains only to linear stochastic systems (classical and quantum); it does not appear in the dissipation inequality for deterministic systems. For details, see [106, 98]

The following theorem relates the property of dissipativeness to certain linear matrix inequalities.

Theorem 8.2.2 ([32]) *Given a quadratic form $r(x, \beta_w)$ defined as above, then the quantum stochastic system (8.1) is dissipative with supply rate $r(x, \beta_w)$ if and only if there exists a real positive definite symmetric matrix X such that the following matrix inequality is satisfied:*

$$\begin{pmatrix} A^T X + XA + R_{11} & R_{12} + XB \\ B^T X + R_{12}^T & R_{22} \end{pmatrix} \leq 0. \tag{8.3}$$

Furthermore, the system is strictly dissipative if and only if there exists a real positive definite symmetric matrix X such that the following matrix inequality is satisfied:

$$\begin{pmatrix} A^T X + XA + R_{11} & R_{12} + XB \\ B^T X + R_{12}^T & R_{22} \end{pmatrix} < 0. \tag{8.4}$$

Moreover, if either of (8.3) or (8.4) holds then the required constant $\lambda \geq 0$ *can be chosen as*

$$\lambda = \mathrm{tr}\left[\left[\begin{array}{c} B^T \\ G^T \end{array}\right] X \left[\begin{array}{cc} B & G \end{array}\right] F\right] \tag{8.5}$$

where the matrix F *is defined by the following relation:*

$$Fdt = \left[\begin{array}{c} dw \\ dv \end{array}\right] \left[\begin{array}{cc} dw^T & dv^T \end{array}\right]. \tag{8.6}$$

We now present some corollaries to the above theorem corresponding to a special case of the matrix R defined in terms of the error output operator

$$\beta_z(t) = Cx(t) + D\beta_w(t).$$

Definition 8.2.3 *The quantum stochastic system (8.1) is said to be* Bounded Real with disturbance attenuation g *if the system (8.1) is dissipative with supply rate*

$$r(x, \beta_w) \;=\; \beta_z^T\beta_z - g^2\beta_w^T\beta_w = [x^T \beta_w^T] \left[\begin{array}{cc} C^TC & C^TD \\ D^TC & D^TD - g^2I \end{array}\right] \left[\begin{array}{c} x \\ \beta_w \end{array}\right].$$

Also, the quantum stochastic system (8.1) is said to be Strictly Bounded Real with disturbance attenuation g *if the system (8.1) is strictly dissipative with this supply rate.*

Using the above definition of a bounded real system, we obtain the following corollary from Theorem 8.2.2 (e.g., see also [107] for the corresponding classical result).

Corollary 8.2.4 ([32]) *The quantum stochastic system (8.1) is bounded real with disturbance attenuation* g *if and only if there exists a positive definite symmetric matrix* $X \in \mathbb{R}^{n\times n}$ *such that the following matrix inequality is satisfied:*

$$\left(\begin{array}{cc} A^TX + XA + C^TC & C^TD + XB \\ B^TX + D^TC & D^TD - g^2I \end{array}\right) \leq 0.$$

Furthermore, the quantum stochastic system is strictly bounded real with disturbance attenuation g *if and only if there exists a positive definite symmetric matrix* $X \in \mathbb{R}^{n\times n}$ *such that the following matrix inequality is satisfied:*

$$\left(\begin{array}{cc} A^TX + XA + C^TC & C^TD + XB \\ B^TX + D^TC & D^TD - g^2I \end{array}\right) < 0.$$

Moreover, in both cases the required constant $\lambda \geq 0$ can be chosen as

$$\lambda = \mathrm{tr}\left[\begin{bmatrix} B^T \\ G^T \end{bmatrix} X \begin{bmatrix} B & G \end{bmatrix} F\right].$$

Now combining this corollary with the standard Strict Bounded Real Lemma (e.g., see [95, 108]) we obtain the following corollary.

Corollary 8.2.5 ([32]) *The following statements are equivalent*

(i) The quantum stochastic system (8.1) is strictly bounded real with disturbance attenuation g.

(ii) A is a stable matrix and $\|C(sI-A)^{-1}B+D\|_\infty < g$.

(iii) $g^2 I - D^T D > 0$ and there exists a positive definite matrix $\tilde{X} > 0$ such that

$$A^T\tilde{X} + \tilde{X}A + C^TC + (\tilde{X}B + C^TD)(g^2I - D^TD)^{-1}(B^T\tilde{X} + D^TC) < 0.$$

(iv) $g^2 I - D^T D > 0$ and the algebraic Riccati equation

$$A^TX + XA + C^TC + (XB + C^TD)(g^2I - D^TD)^{-1}(B^TX + D^TC) = 0$$

has a stabilizing solution $X \geq 0$.

Furthermore, if these statements hold then $X < \tilde{X}$.

Some remarks regarding Corollary 8.2.5 are now in order. It has been shown in [98] that a small gain methodology can be developed for quantum stochastic systems that parallels the small gain methodology in [106] for classical stochastic systems. The essence of the results of [106, 98] is that, properly formulated, the small gain principle (see, e.g., [109]) applies in the same way for linear stochastic systems and an appropriate class of quantum stochastic systems as they do for deterministic systems. In particular, the quantity $\|C(sI-A)^{-1}B+D\|_\infty < g$ in Point 2 of Theorem 8.2.5 implies that:

$$\int_0^t \langle z(s)^T z(s)\rangle ds < g^2 \int_0^t \langle \beta_w(s)^T \beta_w(s)\rangle ds + \mu_1 + \mu_2 t, \quad t \geq 0$$

for some real constants $\mu_1, \mu_2 > 0$. In the terminology of [98], the system (8.1) is then said to be *mean square stable*. The main point of the theorem is that the last property is equivalent to (8.1) being strictly dissipative. Hence mean square

stability is also characterized by strict matrix inequality of Corollary 8.2.4. Now, according to the small gain principle of [98] the smaller the gain g, the more robust the system (8.1) is with respect to some unmodelled dynamics and Wiener noise disturbances which may be present between the signals $z(t)$ and $w(t)$ (see Figure 8.1). In particular, if the gain from $z(t)$ to $w(t)$ due to the unmodelled dynamics and disturbances satisfies:

$$\int_0^t \langle w(s)^T w(s)\rangle ds \leq g'^2 \int_0^t \langle \beta_z(s)^T \beta_z(s)\rangle ds + \mu_1' + \mu_2' t, \quad t \geq 0,$$

for some real constants $\mu_1', \mu_2' > 0$, and $g' \geq 0$ is such that $g'g < 1$ then (8.1) will remain mean square stable. In this case we say that (8.1) is *robustly stable*. Moreover, the smaller the gain g the more robust the system becomes (since this allows a higher gain g' from $z(t)$ to $w(t)$ such that the condition $g'g < 1$ is not violated).

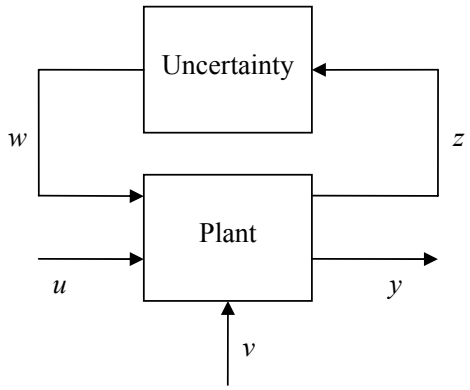

Figure 8.1: The uncertainty block represents unmodelled dynamics and additional quantum Wiener disturbances which may appear between $z(t)$ and $w(t)$

8.3 H^∞ controller synthesis

In this section, we consider the problem of H^∞ controller design for quantum systems. As we shall see, we do not restrict ourselves to classical controllers. The closed loop plant-controller system is defined in Subsection 8.3.1, and then in Subsection 8.3.3 we apply the Strict Bounded Real Lemma to the closed loop system to obtain our main results. In Subsection 8.3.4 we provide conditions under which a controller is physically realizable.

8.3.1 The closed loop plant-controller system

The general linear model (7.2) described above is the prototype for the interconnection of components which will make up the quantum control system. In control system design, we prescribe a system called the *plant*, and seek to find another system, called a *controller*, in such a way that desired closed loop behavior is achieved. We now introduce our plant and controller models, and the resulting closed loop.

We consider *plants* described by non-commutative stochastic models of the following form defined in an analogous way to the quantum system (7.2):

$$\begin{aligned} dx(t) &= Ax(t)dt + [\ B_0 \quad B_1 \quad B_2\ ][\ dv(t)^T \quad dw(t)^T \quad du(t)^T\ ]^T; \quad x(0) = x_0; \\ dz(t) &= C_1 x(t)dt + D_{12}du(t); \\ dy(t) &= C_2 x(t)dt + [\ D_{20} \quad D_{21} \quad 0_{n_y \times n_u}\ ][\ dv(t)^T \quad dw(t)^T \quad du(t)^T\ ]^T. \end{aligned} \tag{8.7}$$

Here $x(t)$ is a vector of plant variables. The input $w(t)$ is represents a *disturbance* signal of the form (7.4). The signal $u(t)$ is a *control* input of the form

$$du(t) = \beta_u(t)dt + d\tilde{u}(t) \tag{8.8}$$

where $\tilde{u}(t)$ is the noise part of $u(t)$ and $\beta_u(t)$ is an adapted, self-adjoint process commuting with $x(t)$. Also, $dv(t)$ represents any additional quantum noise in the plant. The vectors $v(t)$, $\tilde{w}(t)$ and $\tilde{u}(t)$ are independent quantum noises (meaning that they live on distinct Fock spaces) with Ito matrices F_v, $F_{\tilde{w}}$ and $F_{\tilde{u}}$ which are all non-negative Hermitian. We also assume that

$$x(0)x(0)^T - (x(0)x(0)^T)^T = \Theta.$$

The plant is depicted in Figure 8.2.

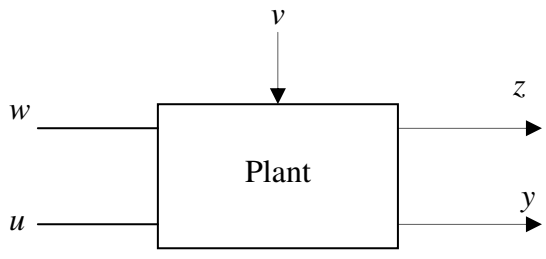

Figure 8.2: Diagram of plant

Controllers are assumed to be non-commutative stochastic systems of the form

$$\begin{aligned} d\xi(t) &= A_K\xi(t)dt + B_{K1}dv_K(t) + B_K dy(t); \quad \xi(0) = \xi_0 \\ du(t) &= C_K\xi(t)dt + B_{K0}dv_K(t) \end{aligned} \tag{8.9}$$

where $\xi(t) = [\ \xi_1(t) \ \ \ldots \ \ \xi_{n_K}(t)\ ]^T$ is a vector of self-adjoint controller variables. The noise $v_K(t) = [\ v_{K1}(t) \ \ \ldots \ \ v_{KK_v}(t)\ ]^T$ is a vector of non-commutative Wiener processes (in vacuum states) with non-zero Ito products as in (7.5) and with canonical Hermitian Ito matrix F_{v_K}, and lives on a distinct Fock space from $v(t)$ and $\tilde{w}(t)$. We will also assume that

$$\xi(0)\xi(0)^T - (\xi(0)\xi(0)^T)^T = \Theta_K.$$

By enlarging the underlying Von Neumann algebra if necessary, the controller can be defined in the above quantum probability space. The controller is depicted in Figure 8.3.

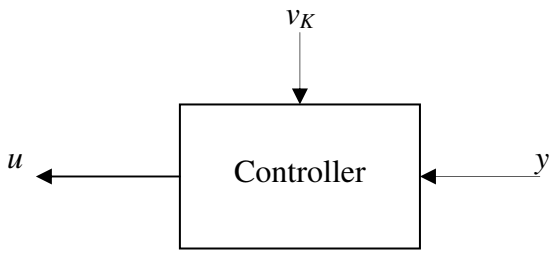

Figure 8.3: Diagram of controller

At time $t = 0$, we also assume that $x(0)$ commutes with $\xi(0)$. The *closed loop system* is obtained by interconnecting (8.7) and (8.9), by identifying $\beta_u(t)$ with $C_K\xi(t)$, to give

$$\begin{aligned} d\eta(t) &= \begin{bmatrix} A & B_2C_K \\ B_KC_2 & A_K \end{bmatrix}\eta(t)dt + \begin{bmatrix} B_0 & B_2B_{K0} \\ B_KD_{20} & B_{K1} \end{bmatrix}\begin{bmatrix} dv(t) \\ dv_K(t) \end{bmatrix} + \\ &\qquad \begin{bmatrix} B_1 \\ B_KD_{21} \end{bmatrix} dw(t); \\ dz(t) &= \begin{bmatrix} C_1 & D_{12}C_K \end{bmatrix}\eta(t)dt + \begin{bmatrix} 0 & D_{12}B_{K0} \end{bmatrix}\begin{bmatrix} dv(t) \\ dv_K(t) \end{bmatrix} \end{aligned} \tag{8.10}$$

where $\eta(t) = [\ x(t)^T \quad \xi(t)^T\ ]^T$. That is, we can write

$$\begin{aligned} d\eta(t) &= \tilde{A}\eta(t)dt + \tilde{B}dw(t) + \tilde{G}d\tilde{v}(t) = \tilde{A}\eta(t)dt + \begin{bmatrix} \tilde{B} & \tilde{G} \end{bmatrix} \begin{bmatrix} dw(t) \\ d\tilde{v}(t) \end{bmatrix}; \\ dz(t) &= \tilde{C}\eta(t)dt + \tilde{H}d\tilde{v}(t) = \tilde{C}\eta(t)dt + \begin{bmatrix} 0 & \tilde{H} \end{bmatrix} \begin{bmatrix} dw(t) \\ d\tilde{v}(t) \end{bmatrix} \end{aligned} \tag{8.11}$$

where

$$\begin{aligned} \tilde{v}(t) &= \begin{bmatrix} v(t) \\ v_K(t) \end{bmatrix}; \quad \tilde{A} = \begin{bmatrix} A & B_2C_K \\ B_KC_2 & A_K \end{bmatrix}; \tilde{B} = \begin{bmatrix} B_1 \\ B_KD_{21} \end{bmatrix}; \\ & \tilde{G} = \begin{bmatrix} B_0 & B_2B_{K0} \\ B_KD_{20} & B_{K1} \end{bmatrix}; \\ \tilde{C} &= \begin{bmatrix} C_1 & D_{12}C_K \end{bmatrix}; \quad \tilde{H} = \begin{bmatrix} 0 & D_{12}B_{K0} \end{bmatrix}. \end{aligned}$$

Note that the closed loop system (8.11) is a system of the form (7.2). It is depicted in Figure 8.4.

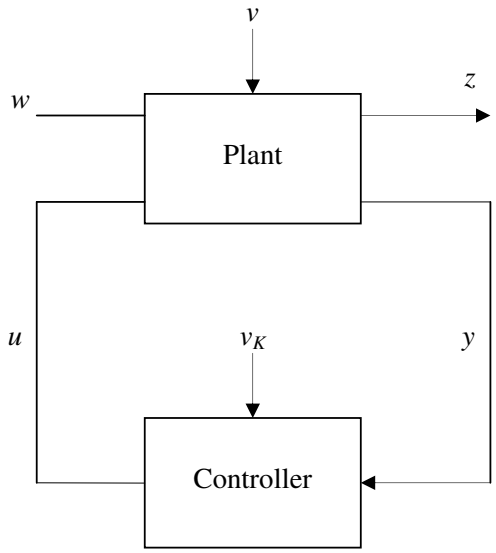

Figure 8.4: Diagram of closed loop system

Remark 8.3.1 *An important aspect to be noted about the way we have defined our plant and controller models is that their outputs are specified such that there is no direct feedthrough of their respective input signals. That is, in the plant $u(t)$ only influences $y(t)$ indirectly via $x(t)$, likewise in the controller $y(t)$ only*

influences $u(t)$ indirectly via $\xi(t)$. This avoids delicate and difficult physical issues that arise when there is direct feedthrough, due to the self-interaction of both the plant and controller in the feedback loop (for example, see [86, Appendix II]).

8.3.2 H^∞ control objective

The goal of the H^∞ controller synthesis is to find a controller (8.9) such that for a *given* disturbance attenuation parameter $g > 0$:

$$\int_0^t \langle z(s)^T z(s) \rangle ds < g^2 \int_0^t \langle \beta_w(s)^T \beta_w(s) \rangle ds + \mu_1 + \mu_2 t,$$

is satisfied for some real constants $\mu_1, \mu_2 > 0$. This objective can be interpreted as that of *disturbance attenuation* where the controller bounds the effect of the "energy" of the signal $\beta_w(t)$ and the noise variances on the "energy" of the signal $z(t)$. Consequently, as explained at the end of Section 8.2, the controller is *robustly stabilizing*. Naturally, one would like to have g as small as possible, but if it is too small a desired controller may not exist. Necessary and sufficient conditions for the existence of a specific type of controller which achieves this goal for a given g are given in the next section, as well as explicit formulas for A_K, B_K and C_K. The results parallel the corresponding well-known results for classical linear systems (see, e.g., [110, 95]).

8.3.3 Necessary and sufficient conditions

In order to present the results on quantum H^∞ control, we will require that the plant system (8.7) satisfies the following assumptions.

Assumption 8.3.2

1. $D_{12}^T D_{12} = E_1 > 0$.
2. $D_{21} D_{21}^T = E_2 > 0$.
3. *The matrix* $\begin{bmatrix} A - j\omega I & B_2 \\ C_1 & D_{12} \end{bmatrix}$ *is full rank for all* $\omega \geq 0$.
4. *The matrix* $\begin{bmatrix} A - j\omega I & B_1 \\ C_2 & D_{21} \end{bmatrix}$ *is full rank for all* $\omega \geq 0$.

The results will be stated in terms of the following pair of algebraic Riccati equations:

$$(A - B_2E_1^{-1}D_{12}^TC_1)^TX + X(A - B_2E_1^{-1}D_{12}^TC_1) + X(B_1B_1^T - g^2B_2E_1^{-1}B_2')X + g^{-2}C_1^T(I - D_{12}E_1^{-1}D_{12}^T)C_1 = 0; \quad (8.12)$$

$$(A - B_1D_{21}^TE_2^{-1}C_2)Y + Y(A - B_1D_{21}^TE_2^{-1}C_2) + Y(g^{-2}C_1^TC_1 - C_2^TE_2^{-1}C_2)Y + B_1(I - D_{21}^TE_2^{-1}D_{21})B_1^T = 0. \quad (8.13)$$

The solutions to these Riccati equations will be required to satisfy the following assumption.

Assumption 8.3.3

(i) $A - B_2E_1^{-1}D_{12}^TC_1 + (B_1B_1^T - g^2B_2E_1^{-1}B_2')X$ *is a stability matrix.*

(ii) $A - B_1D_{21}^TE_2^{-1}C_2 + Y(g^{-2}C_1^TC_1 - C_2^TE_2^{-1}C_2)$ *is a stability matrix.*

(iii) The matrix XY *has a spectral radius strictly less than one.*

It will be shown that if the Riccati equations (8.12), (8.13) have solutions satisfying Assumption 8.3.3, then a controller of the form (8.9) will solve the H^∞ control problem under consideration if its system matrices are constructed from the Riccati solutions as follows:

$$\begin{aligned} A_K &= A + B_2C_K - B_KC_2 + (B_1 - B_KD_{21})B_1^TX; \\ B_K &= (I - YX)^{-1}(YC_2^T + B_1D_{21}^T)E_2^{-1}; \\ C_K &= -E_1^{-1}(g^2B_2^TX + D_{12}^TC_1). \end{aligned} \quad (8.14)$$

We are now in a position to present the main result in [32] concerning H^∞ controller synthesis which follows directly from the classical results in [110, 95].

Theorem 8.3.4 ([32]) (Necessity) *Consider the system (8.7) and suppose that Assumption 8.3.2 is satisfied. If there exists a controller of the form (8.9) such that the resulting closed loop system (8.11) is strictly bounded real with disturbance attenuation* g*, then the Riccati equations (8.12), (8.13) will have stabilizing solutions* $X \geq 0$ *and* $Y \geq 0$ *satisfying Assumption 8.3.3.*

(Sufficiency) *Suppose the Riccati equations (8.12), (8.13) have stabilizing solutions* $X \geq 0$ *and* $Y \geq 0$ *satisfying Assumption 8.3.3. If the controller (8.9)*

is such that the matrices A_K, B_K, C_K are as defined in (8.14), then the resulting closed loop system (8.11) will be strictly bounded real with disturbance attenuation g. Also the constant $\lambda \geq 0$ in Definition 8.2.1 can be chosen as in (8.5) with B, G, and F replaced by $\tilde{B}$, $\tilde{G}$ and $\tilde{F}$, where $\tilde{F}$ is defined by the relation:

$$\tilde{F}dt = \begin{bmatrix} dw(t) \\ d\tilde{v}(t) \end{bmatrix} \begin{bmatrix} dw(t)^T & d\tilde{v}(t)^T \end{bmatrix}$$

Notice that the controller parameters B_{K0}, B_{K1}, and the controller noise v_K are not given in the construction described in the sufficiency part of Theorem 8.3.4. In fact, they are free as far as the H^∞ objective is concerned. In the next subsection, we show that they may always be chosen to yield a physically realizable controller.

8.3.4 Physical realization of controllers

In this section we will show that given an *arbitrary* choice of a commutation matrix $\Theta_K = \xi(0)\xi(0)^T - (\xi(0)\xi(0)^T)^T$ for the controller, there always exists a physically realizable controller in the sense of Definition 7.3.5. This is a rather surprising result since it implies that the controller can be chosen to be purely quantum, purely classical, or a combination of quantum and classical components at will.

Theorem 8.3.5 *Assume*

$$F_y = D_{20}F_vD_{20}^T + D_{21}F_wD_{21}^T$$

is canonical. Let $\{A_K, B_K, C_K\}$ be an arbitrary triple (such as given by (8.14)), and select the controller commutation matrix Θ_K to be canonical or degenerate canonical, as desired. Then there exists controller parameters B_{K0}, B_{K1}, and the controller noise v_K such that the controller (8.9) is physically realizable. In particular, $2i\Theta_K = \xi(t)\xi(t)^T - (\xi(t)\xi(t)^T)^T$ for all $t \geq 0$ whenever $2i\Theta_K = \xi(0)\xi(0)^T - (\xi(0)\xi(0)^T)^T$.

The proof of this theorem depends on the following lemma for the case in which Θ_K is canonical. For the degenerate canonical case, this lemma can be applied to an augmentation of the controller. We shall use the notation of Section 7.3.1, and as in the discussion in Section 7.2, we may take B_K to have an even number of columns and C_K to have an even number of rows.

Lemma 8.3.6 *Let F_y be canonical and $\{A_K, B_K, C_K\}$ be such that $A_K \in \mathbb{R}^{n_K \times n_K}$, $B_K \in \mathbb{R}^{n_K \times m_K}$, $C_K \in \mathbb{R}^{l_k \times n_K}$, $n_K = 2N_\xi$, $m_K = 2N_y$ and $l_K = 2N_u$ for positive integers N_ξ, N_y and N_u, and $\Theta_K = \mathrm{diag}_{N_\xi}(J)$ is canonical. Then there exists an integer $N_{v_K} \geq N_u$ and $B_{K1} \in \mathbb{R}^{n_K \times 2N_{w_K}}$, with $N_{w_K} = N_{v_K} + N_y$, such that the system (8.9) is physically realizable with*

$$B_{K0} = P_{N_u}^T \begin{bmatrix} \Sigma_{N_u} & 0_{N_u \times N_{w_K}} \\ 0_{N_u \times N_{w_K}} & \Sigma_{N_u} \end{bmatrix} P_{N_{w_K}} \begin{bmatrix} I_{2N_{v_K} \times 2N_{v_K}} \\ 0_{m_k \times 2N_{v_K}} \end{bmatrix} = \begin{bmatrix} I_{n_y \times n_y} & 0_{n_y \times (n_w - n_y)} \end{bmatrix},$$

$$R = \frac{1}{2}(Z + Z^T); \tag{8.15}$$

$$B_{K1} = \begin{bmatrix} B_{K1,1} & B_{K1,2} \end{bmatrix}; \tag{8.16}$$

$$\Lambda = \begin{bmatrix} \frac{1}{2} C_K^T P_{N_u}^T \begin{bmatrix} I \\ iI \end{bmatrix} & \Lambda_{b1}^T & \Lambda_{b2}^T \end{bmatrix}^T; \tag{8.17}$$

$$B_{K1,1} = -i\Theta_K C_K^T \mathrm{diag}_{N_u}(iJ); \tag{8.18}$$

$$\Lambda_{b2} = -i \begin{bmatrix} I_{N_y \times N_y} & 0_{N_y \times N_y} \end{bmatrix} P_{N_y} \mathrm{diag}_{N_y}(M) B_K^T \Theta_K; \tag{8.19}$$

$$B_{K1,2} = 2i\Theta_K \begin{bmatrix} -\Lambda_{b1}^\dagger & \Lambda_{b1}^T \end{bmatrix} P_{N_{v_K} - N_u} \mathrm{diag}_{N_{v_K} - N_u}(M) \tag{8.20}$$

where $Z = -\frac{1}{2}\Theta A_K$ and $N_{v_K} \geq N_u + 1$. Here Λ_{b1} is any complex $(N_{v_K} - N_u) \times n_K$ matrix such that

$$\Lambda_{b1}^\dagger \Lambda_{b1} = \Xi + i \left(\frac{1}{2}(Z - Z^T) - \frac{1}{4} C_K^T P_{N_u}^T \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P_{N_u} C_K - \Im(\Lambda_{b2}^\dagger \Lambda_{b2}) \right), \tag{8.21}$$

where Ξ is any real symmetric $n_K \times n_K$ matrix such that the right hand side of (8.21) is non-negative definite.

Remark 8.3.7 *Note that the condition $N_{v_K} \geq N_u$ is significant since it implies that there is no direct feedthrough of the signal $y(t)$ to $u(t)$ (see Remark 8.3.1) as required for (8.9). For compatibility between the equations (8.9) and (8.7), it is necessary that the corresponding Ito matrices satisfy the following condition:*

$$F_u = B_{K0} F_{v_K} B_{K0}^T. \tag{8.22}$$

However, since F_{v_K} and F_u are, by convention, in canonical form, (8.22) is always satisfied. To see this, we simply note that the $2N_u$ elements of $B_{K0}v_K$ are a subset of pairs of conjugate real and imaginary quadratures in v_K. Hence it follows that if F_{v_K} is canonical then F_u must also be canonical and (8.22) is automatically satisfied.

The proof of Lemma 8.3.5 uses the following lemma:

Lemma 8.3.8 *If S is a Hermitian matrix then there is a real constant α_0 such that $\alpha I + S \geq 0$ for all $\alpha \geq \alpha_0$.*

Proof. Since S is Hermitian it has real eigenvalues and is diagonalizable. Hence $S = V^\dagger E V$ for some real diagonal matrix E and unitary matrix V. Now let $\alpha_0 = -\lambda$, where λ is the smallest eigenvalue of S. The result follows since $\alpha I + S = V^\dagger(\alpha I + E)V$ while $\alpha I + E \geq 0$ for all $\alpha \geq \alpha_0$. □

Proof. (of Lemma 8.3.5) The main idea is to explicitly construct matrices $R \in \mathbb{R}^{n_K \times n_K}$, $\Lambda \in \mathbb{C}^{N_{v_K} \times n_K}$, $B_{K1} \in \mathbb{R}^{n_K \times 2(N_{v_K}+N_y)}$ and $B_{K0} \in \mathbb{R}^{l_K \times 2N_{v_K}}$, with $N_{v_K} \geq N_u$, such that (7.14)-(7.17) are satisfied by identifying A_K, B_K, C_K, $[\ B_{K0}\ \ 0_{l_K \times m_K}\ ]$, ξ, w_K and u with A, B, C, D, x, w and y, respectively. To this end, let $Z = \frac{1}{2}\Theta_K^{-1}A = -\frac{1}{2}\Theta_K A$, with $\Theta_K = \mathrm{diag}_{N_\xi}(J)$. We first construct matrices Λ_{b2}, Λ_{b1}, $B_{K1,1}$ and $B_{K1,2}$ according to the following procedure:

1. Construct the matrix Λ_{b2} according to (8.19).

2. Construct a real symmetric $n_K \times n_K$ matrix Ξ_1 such that the matrix

$$\Xi_2 \;=\; \Xi_1 + i\left(\frac{Z - Z^T}{2} - \frac{1}{4}C_K^T P_{N_u}^T \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P_{N_u} C_K - \Im(\Lambda_{b2}^\dagger \Lambda_{b2})\right)$$

   is non-negative definite. It follows from Lemma 8.3.8 that such a matrix Ξ_1 always exists.

3. Construct a matrix Λ_{b1} such that $\Lambda_{b1}^\dagger \Lambda_{b1} = \Xi_2$. This can be done, for example, using the singular value decomposition of Ξ_2 (in this case Λ_{b1} will have n_K rows).

4. Construct the matrices $B_{K1,1}$ and $B_{K1,2}$ according to equations (8.18) and (8.20), respectively.

Let $R = \frac{1}{2}(Z + Z^T)$. We now show that there exists an integer $N_q^{v_K} \geq N_q^u$ such conditions (7.14)-(7.17) are satisfied with the matrix R as defined and with $B_{K1} = [\ B_{K1,1} \quad B_{K1,2}\ ]$ and

$$\Lambda = \begin{bmatrix} \frac{1}{2}\begin{bmatrix} I & iI \end{bmatrix} P_{N_u} C_k \\ \Lambda_{b1} \\ \Lambda_{b2} \end{bmatrix}. \tag{8.23}$$

First note that necessarily $N_{v_K} \geq N_u + 1 > N_u$ since B_{K1} has at least $2N_u + 2$ columns. Also, by virtue of our choice of Λ_{b1} we have

$$\Im(\Lambda_{b1}^\dagger \Lambda_{b1}) = \Im(\Xi_2) = \frac{1}{2}(Z - Z^T) - \frac{1}{4} C_K^T P_{N_u}^T \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P_{N_u} C_K - \Im(\Lambda_{b2}^\dagger \Lambda_{b2}),$$

and hence

$$\Im(\Lambda^\dagger \Lambda) = \Im(\Lambda_{b1}^\dagger \Lambda_{b1}) + \Im(\Lambda_{b2}^\dagger \Lambda_{b2}) + \frac{1}{4} C_K^T P_{N_u}^T \begin{bmatrix} 0 & I \\ -I & 0 \end{bmatrix} P_{N_u} C_K = \frac{1}{2}(Z - Z^T).$$

Since $R = \frac{Z+Z^T}{2}$, we have $R + \Im(\Lambda^\dagger \Lambda) = Z$. Therefore, (7.14) is satisfied.

Now, as in the proof of Theorem 7.3.6, observe that

$$i\Theta_K B_K \mathrm{diag}_{N_y}(M^\dagger) P_{N_y}^T = [\ T \quad -T^{\#}\ ]$$

for some $n_K \times N_y$ complex matrix T. But by taking the conjugate transpose of both sides of (8.19) which defined Λ_{b2}, we conclude that $T = -\Lambda_{b2}^\dagger$. Hence,

$$B_K = 2i\Theta_K [\ -\Lambda_{b2}^\dagger \quad \Lambda_{b2}^T\ ] P_{N_y} \mathrm{diag}_{N_y}(M). \tag{8.24}$$

From (8.18) which defined Λ_{b1}, we obtain

$$\begin{aligned} B_{K1,1} &= -i\Theta_K C_K^T \mathrm{diag}_{N_u}(iJ) \\ &= -i\Theta_K C_K^T \mathrm{diag}_{N_u}(iJ)(2\mathrm{diag}_{N_u}(M^\dagger))\mathrm{diag}_{N_u}(M) \\ &= i\Theta_K C_K^T \mathrm{diag}_{N_u}\left(\begin{bmatrix} -1 & 1 \\ i & i \end{bmatrix}\right)\mathrm{diag}_{N_u}(M) \\ &= i\Theta_K C_K^T P_{N_u}^T \begin{bmatrix} -I & I \\ iI & iI \end{bmatrix} P_{N_u} \mathrm{diag}_{N_u}(M). \end{aligned} \tag{8.25}$$

Combining (8.20), (8.24) and (8.25) gives us

$$
\begin{aligned}
&\left[\begin{array}{ccc} B_{K1,1} & B_{K1,2} & B_K \end{array}\right] \\
&= 2i\Theta_K \left[\tfrac{1}{2} C_K^T P_{N_u}^T \left[\begin{array}{cc} -I & I \\ iI & iI \end{array}\right] P_{N_u} \left[\begin{array}{cc} -\Lambda_{b1}^\dagger & \Lambda_{b1}^T \end{array}\right] P_{(N_{v_K}-N_u)} \right. \\
&\qquad \left. \left[\begin{array}{cc} -\Lambda_{b2}^\dagger & \Lambda_{b2}^T \end{array}\right] P_{N_y} \right] P_{N_{w_K}}^T P_{N_{w_K}} \mathrm{diag}_{n_{w_K}}(M) \\
&= 2i\Theta_K \left[-\tfrac{1}{2} C_K^T P_{N_u}^T \left[\begin{array}{c} I \\ -iI \end{array}\right] \quad -\Lambda_{b1}^\dagger \quad -\Lambda_{b2}^\dagger \quad \tfrac{1}{2} C_K^T P_{N_u}^T \left[\begin{array}{c} I \\ iI \end{array}\right] \quad \Lambda_{b1}^T \quad \Lambda_{b2}^T \right] \\
&\qquad P_{N_{w_K}} \mathrm{diag}_{N_{w_K}}(M) \\
&= 2i\Theta_K \left[-\tfrac{1}{2} C_K^T P_{N_u}^T \left[\begin{array}{c} I \\ -iI \end{array}\right] \quad -\Lambda_{b1}^\dagger \quad -\Lambda_{b2}^\dagger \quad \tfrac{1}{2} C_K^T P_{N_u}^T \left[\begin{array}{c} I \\ iI \end{array}\right] \quad \Lambda_{b1}^T \quad \Lambda_{b2}^T \right] \Gamma \\
&= 2i\Theta_K \left[\begin{array}{cc} -\Lambda^\dagger & \Lambda^T \end{array}\right] \Gamma.
\end{aligned}
$$

Therefore, (7.15) is also satisfied. Moreover, it is straightforward to verify (7.16) by substituting Λ as defined by (8.23) into the right hand side of (7.16). Finally, since $N_{v_K} > N_u$, it follows that $\left[\begin{array}{cc} B_{K0} & 0_{l_K \times m_k} \end{array}\right]$ is precisely the right hand side of (7.17). This completes the proof of Lemma 8.3.5. □

8.4 H^∞ synthesis in quantum optics

Quantum optics is an important area in quantum physics and quantum technology and provides a promising means of implementing quantum information and computing devices; e.g., see [90]. In this section we give some examples of controller design for simple quantum optics plants based on optical cavities and optical amplifiers coupled to optical fields; e.g., see [111, 93]. We give explicit realizations of controllers which are fully quantum, fully classical, and mixed quantum-classical using standard quantum optical components and electronics.

8.4.1 Quantum controller synthesis

We consider an optical cavity resonantly coupled to three optical channels v, w, u as in Figure 8.5.

The dynamics of this cavity system is described by the evolution of its annihilation operator a (representing a quantized single mode standing wave). In

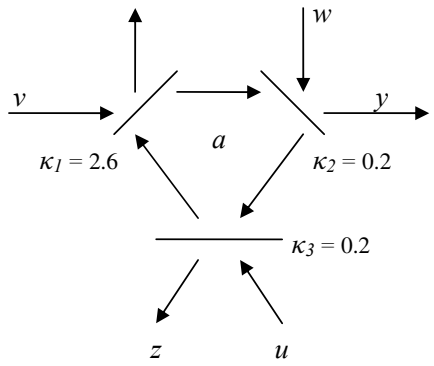

Figure 8.5: An optical cavity (plant).

the quadrature notation of (8.7), $x_1(t) = q(t) = a(t) + a^*(t)$, $x_2(t) = p(t) = (a(t) - a^*(t)/i$, $v(t) = (v_1(t), v_2(t))^T$, $w(t) = (w_1(t), w_2(t))^T$, $u(t) = (u_1(t), u_2(t))^T$. The quantum noises v, $\tilde{w}$ have Hermitian Ito matrices $F_v = F_{\tilde{w}} = I + iJ$. This leads to a system of the form (8.7) with the following system matrices:

$$\begin{aligned} A &= -\frac{\gamma}{2}I; \quad B_0 = -\sqrt{\kappa_1}I; \quad B_1 = -\sqrt{\kappa_2}I; \quad B_2 = -\sqrt{\kappa_3}I; \\ &\quad (\gamma = \kappa_1 + \kappa_2 + \kappa_3) \\ C_1 &= \sqrt{\kappa_3}I; \quad D_{12} = I; \\ C_2 &= \sqrt{\kappa_2}I; \quad D_{21} = I. \end{aligned} \tag{8.26}$$

In this model, the boson commutation relation $[a, a^*] = 1$ holds. This means that the commutation matrix for this plant is $\Theta_P = J$.

In our example, we will choose the total cavity decay rate $\kappa = 3$ and the coupling coefficients $\kappa_1 = 2.6$, $\kappa_2 = \kappa_3 = 0.2$. With a disturbance attenuation constant of $g = 0.1$, it was found that the Riccati equations (8.12) and (8.13) have stabilizing solutions satisfying Assumption 8.3.3. These Riccati solutions were as follows: $X = Y = 0_{2\times 2}$. Then, it follows from Theorem 8.2.2 that if a controller of the form (8.9) is applied to this system with matrices A_K, B_K, C_K defined as in (8.14) then the resulting closed loop system will be strictly bounded real with disturbance attenuation g. In our case, these matrices are given by

$$A_K = -1.1I, \; B_K = -0.447I, \; C_K = -0.447I.$$

The form of the matrices A_K, B_K and C_K suggest that the controller (8.9) can be implemented using an optical cavity in combination with some phase shifters. Indeed, it is a straightforward exercise to devise an implementation as shown in Figure 8.6.

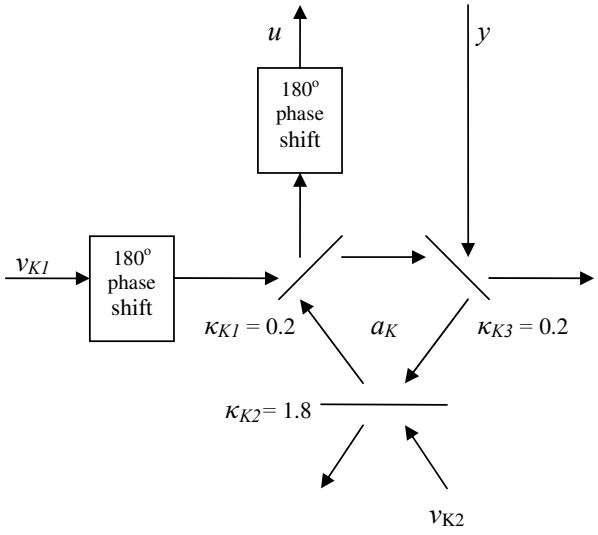

Figure 8.6: An optical cavity quantum realization of the controller ($\Theta_K = J$) for the plant shown in Figure 8.5.

In the figure, a_K is the annihilation operator of the cavity, $\xi_1 = a_K + a_K^*$, $\xi_2 = (a_K - a_K^*)/i$ and $\xi = (\xi_1, \xi_2)^T$. The implemented controller is a physically realizable system with the following dynamics:

$$\begin{aligned} d\xi(t) &= A_K\xi(t)dt + [\ B_{K1} \quad B_K\ ][\ dv_K^T \quad dy^T\ ]^T \\ d\tilde{u}(t) &= C_K\xi(t)dt + [\ I_{2\times 2} \quad 0_{2\times 4}\ ][\ dv_K^T \quad dy^T\ ]^T, \end{aligned} \tag{8.27}$$

where $B_{K1} = [\ \sqrt{\kappa_{K1}}I_{2\times 2} \quad -\sqrt{\kappa_{K2}}I_{2\times 2}\ ]$ and $v_K = [\ v_{K1}^T \quad v_{K2}^T\ ]^T$, with $\kappa_{K1} = 0.2$ and $\kappa_{K2} = 1.8$. For this realization we have

$$R = 0_{2\times 2} \quad \Lambda = \begin{bmatrix} -0.2236 & -0.2236i \\ 0.6708 & 0.6708i \\ 0.2235 & 0.2235i \end{bmatrix}.$$

Note that the controller of Figure 8.6 requires two phase shifters. Alternatively, if we consider realizing the controller as a cascade of a passive optical device with a physically realizable system, see the discussion in Remark 7.3.8, it is possible to remove one of the phase shifters. This alternative realization is illustrated in Figure 8.7.

Now the controller is implemented as an optical cavity, with annihilation operator a_K, connected at the output with a 180° phase shifter. The controller cavity has coupling coefficients $\kappa_{K1} = 0.2$, $\kappa_{K2} = 1.8$, $\kappa_{K3} = 0.2$ and is a physically

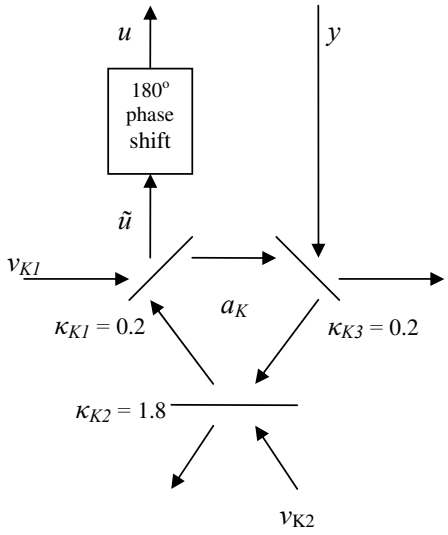

Figure 8.7: An alternative optical cavity quantum realization of the controller ($\Theta_K = J$) for the plant shown in Figure 8.5.

realizable system with dynamics:

$$\begin{aligned} d\xi(t) &= A_K\xi(t)dt + [\ B_{K1} \quad B_K\ ][\ dv_K^T \quad dy^T\ ]^T \\ d\tilde{u}(t) &= -C_K\xi(t)dt + [\ I_{2\times 2} \quad 0_{2\times 4}\ ][\ dv_K^T \quad dy^T\ ]^T, \end{aligned} \tag{8.28}$$

where $\tilde{u}(t)$ is the output of the cavity, $B_{K1} = [\ -\sqrt{\kappa_{K1}}I_{2\times 2} \quad -\sqrt{\kappa_{K2}}I_{2\times 2}\ ]$ and $v_K = [\ v_{K1}^T \quad v_{K2}^T\ ]^T$. For the cavity we have that

$$R = 0_{2\times 2} \quad \Lambda = \begin{bmatrix} 0.2236 & 0.2236i \\ 0.6708 & 0.6708i \\ 0.2235 & 0.2235i \end{bmatrix}.$$

The overall output of the controller is $u(t)$, given by $u(t) = K_s\tilde{u}(t)$, where $K_s = -I_{2\times 2}$. Here K_s models the 180° phase shift at the output of the cavity (cf. Remark 7.3.8). Thus, the overall controller is of the form (8.9) with $B_{K0} = [-\ I \quad 0\ ]$ and B_{K1} as given before.

Notice that the B_{K1} matrix in (8.27) and the B_{K1} matrix in (8.28) differ in the sign of the $(1,1)$ 2×2 block, which is why the second realization does not require a phase shifter in front of v_{K1}.

8.4.2 Classical controller synthesis

In subsection 8.4.1 we obtained a quantum controller corresponding to the choice $\Theta_K = J$. We now show that if we instead choose $\Theta_K = 0$, the controller that is

realized is classical, with appropriate transitions to and from the quantum plant.

Now, suppose we choose v_K to be the quadratures of two independent noise channels (i.e., $F_{v_K} = I_{4\times 4} + i\mathrm{diag}(J, J)$). Setting $\Theta_K = 0_{2\times 2}$, Eq. (7.19) and the compatibility requirement (8.22) in this context results in the following pair of equations:

$$B_K J B_K^T + B_{K1}\mathrm{diag}(J, J)B_{K1}^T = 0 \tag{8.29}$$

$$B_{K0}(I_{4\times 4} + i\mathrm{diag}(J, J))B_{K0}^T = I + iJ. \tag{8.30}$$

In order to find B_{K0} and B_{K1} solving (8.29) and (8.30), we assume the following forms for B_{K0} and B_{K1}:

$$B_{K0} = \left[\begin{array}{cc} \tilde{B}_{K0} & 0_{2\times 2} \end{array}\right]; \quad B_{K1} = \left[\begin{array}{cc} 0_{2\times 2} & \tilde{B}_{K1} \end{array}\right].$$

Since $B_K = -0.447I$, substitution of these forms into (8.29) and (8.30) gives:

$$\tilde{B}_{K0}(I + iJ)\tilde{B}_{K_0}^T = I + iJ; \quad 0.447^2 J + \tilde{B}_{K1} J \tilde{B}_{K1}^T = 0.$$

It can be readily checked, by direct substitution, that these equations are solved by $\tilde{B}_{K0} = I_{2\times 2}$ and $\tilde{B}_{K1} = -0.447\tilde{I}$, where $\tilde{I} = \left[\begin{array}{cc} 1 & 0 \\ 0 & -1 \end{array}\right]$. This completely specifies the classical realization of the controller, illustrated in Figure 8.8. The quantum signal y is converted to a classical signal $y_c = (y_{c1}, y_{c2})^T = (y_1 - v_{K21}, v_{K22} + y_2)^T$ by imperfect continuous measurement of the real and imaginary quadratures of the optical beam, implemented in Figure 8.8 by a beam splitter and two homodyne detectors [111]. The classical signal y_c is processed by a classical linear system $(A_K, B_K, C_K, 0)$ to produce a classical control signal u_c, which then modulates (displaces) a field v_{K1} to produce the optical control signal $du = u_c dt + dv_{K1}$. This classical controller achieves exactly the same H^∞ performance as the quantum controller of Subsection 8.4.1.

This classical controller has access to the full quantum signal y, and the quantum measurement occurs in the controller. The algebra based on the commutation relations enforces the quantum measurement, and also the modulation. If we were to include measurement as part of the plant specification, then in general a different classical controller will result, with different H^∞ performance. To see this, suppose that y is replaced by its real quadrature in the plant specification;

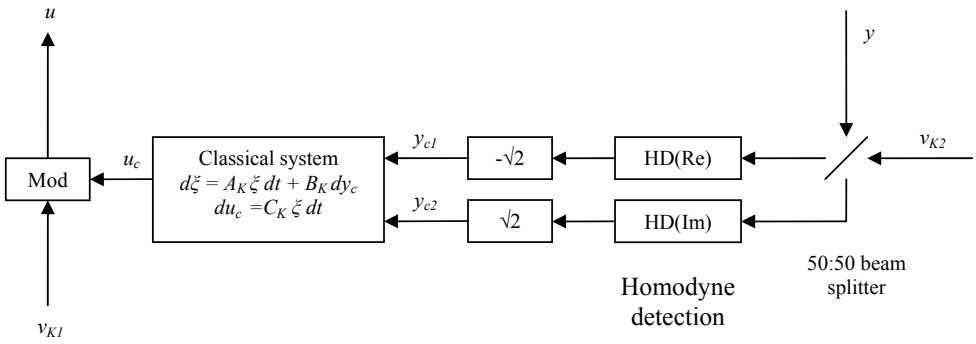

Figure 8.8: A classical realization of the controller ($\Theta_K = 0$) for the plant shown in Figure 8.5. The controller includes quantum measurement and classical modulation of optical fields.

this situation is described by the matrices

$$\begin{aligned} A &= -\frac{\gamma}{2} I; \quad B_0 = -\sqrt{\kappa_1} I; \quad B_1 = -\sqrt{\kappa_2} I; \quad B_2 = -\sqrt{\kappa_3} I; \\ &\quad (\gamma = \kappa_1 + \kappa_2 + \kappa_3) \\ C_1 &= \sqrt{\kappa_3} I; \quad D_{12} = I; \\ C_2 &= \sqrt{\kappa_2} \begin{bmatrix} 1 & 0 \end{bmatrix}; \quad D_{21} = \begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned} \tag{8.31}$$

and is illustrated in Figure 8.9. Thus the output of the plant is a classical single-variable signal.

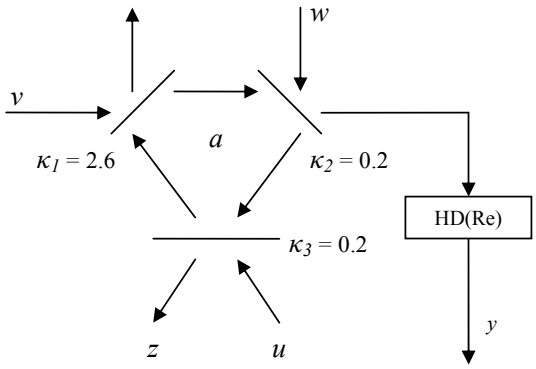

Figure 8.9: An optical cavity (plant) with classical output. The (real) quadrature measurement is achieved by homodyne photodetection (HD(Re)).

With a disturbance attenuation constant of $g = 0.134$, it was found that the Riccati equations (8.12) and (8.13) have the following stabilizing solutions

satisfying Assumption 8.3.3:

$$X = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}; \quad Y = \begin{bmatrix} 0 & 0 \\ 0 & 0.121 \end{bmatrix}.$$

It now follows from Theorem 8.2.2 that if a controller of the form (8.9) is applied to this system with the following matrices A_K, B_K, C_K defined as in (8.14), then the resulting closed loop system will be strictly bounded real with disturbance attenuation $g = 0.134$:

$$\begin{aligned} A_K &= \begin{bmatrix} -1.1 & 0 \\ 0 & -1.3 \end{bmatrix}; \quad B_K = \begin{bmatrix} -0.447 \\ 0 \end{bmatrix}; \\ C_K &= \begin{bmatrix} -0.447 & 0 \\ 0 & -0.447 \end{bmatrix}. \end{aligned} \tag{8.32}$$

In this case, the controller (8.9), (8.32) is a classical system which can be implemented using standard electronic devices. This second classical controller is illustrated in Figure 8.10, and is different to the previous one. Here we have chosen $B_{K0} = I$, $B_{K1} = 0$, and the quantum noise is canonical. The control signal is $du = u_c dt + dv_K$, a coherent optical field.

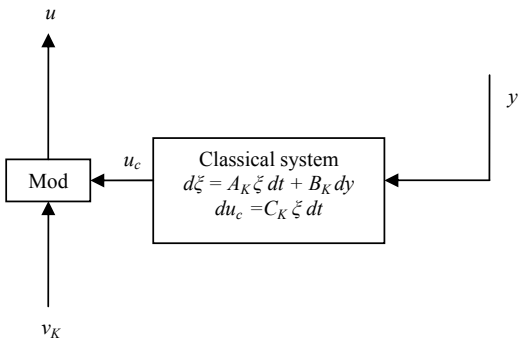

Figure 8.10: Classical controller ($\Theta_K = 0$) for the plant of Figure 8.9.

Finally, let us now return to the controller of Figure 8.8, which we had constructed "directly", and show how it may be recovered using Lemma 8.3.6. To this end, suppose we seek a physically realizable controller of the form:

$$\begin{aligned} d\xi(t) &= A_K \xi(t) dt + B_{K1} dv_K(t) + B_K dy(t) \\ du(t) &= C_K \xi(t) dt + B_{K0} dv_K(t), \end{aligned} \tag{8.33}$$

with $\Theta_K = 0$, and the real matrices B_{K0} and B_{K1} are unknown and to be determined. Since Θ_K is degenerate canonical, we need to consider some augmentation of (8.33) (Theorem 8.3.5). Suppose we consider some augmentation where A_K, B_K, C_K are, respectively, augmented to the matrices $\tilde{A}_K, \tilde{B}_K, \tilde{C}_K$ given by:

$$\tilde{A}_K = \begin{bmatrix} & A_K & & 0_{2\times 2} \\ 0 & & -0.447^2/2 & \\ & -A_K^T & & \\ 0.447^2/2 & & 0 & \end{bmatrix} \quad \tilde{B}_K = \begin{bmatrix} B_K \\ 0_{2\times 2} \end{bmatrix}$$
$$\tilde{C}_K = \begin{bmatrix} C_K & 0_{2\times 2} \end{bmatrix},$$

and let P be a permutation matrix such that $P^T\tilde{\Theta}_K P^T$ is canonical. Now, applying Lemma 8.3.6 to the triplet $\{P\tilde{A}_K P^T, P\tilde{B}_K, \tilde{C}_K P^T\}$ by choosing

$$\Xi = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0.05 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.05 \end{bmatrix} \text{ and } \Lambda_{b1} = \begin{bmatrix} 0 & 0.2235i & 0 & 0.2235 \end{bmatrix},$$

gives us a physically realizable dilation of (8.33) for some B_{K0} and B_{K1}. By extracting the sub-system of the dilation corresponding to our original system (8.33), we recover the matrices $B_{K0} = \begin{bmatrix} I_{2\times 2} & 0_{2\times 2} \end{bmatrix}$ and $B_{K1} = \begin{bmatrix} 0 & -0.447\tilde{I} \end{bmatrix}$ as we had before. This results in the controller given in Figure 8.8.

8.4.3 Classical-quantum controller synthesis

As a final example, we illustrate the synthesis of a controller with both classical and quantum components. The plant has two degrees of freedom, and is formed as a cascade of an optical amplifier [93] and the cavity discussed above. This plant is illustrated in Figure 8.11.

The optical amplifier has an auxiliary input h, which is an inverted heat bath with Ito matrix $F_h = (2N+1)I + iJ$, where $N > 0$ is a positive thermal parameter (for details, see [93]). The complete system shown in Figure 8.11 is of the form

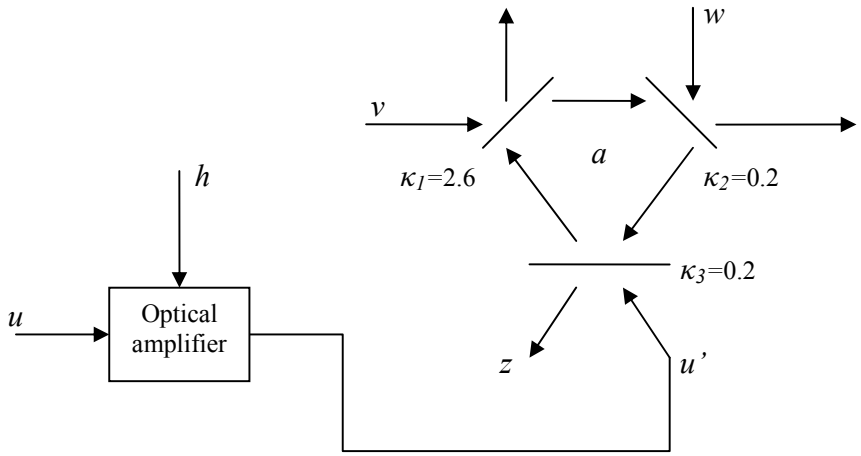

Figure 8.11: An optical amplifier-cavity system (plant).

(8.7) with matrices

$$
\begin{aligned}
A &= \begin{bmatrix} -\frac{\gamma}{2}I & -\sqrt{\kappa_3\alpha}\,I \\ 0 & -\frac{\alpha-\beta}{2}I \end{bmatrix};\quad B_0 = \begin{bmatrix} -\sqrt{\kappa_1}I & 0 \\ 0 & \sqrt{\beta}I \end{bmatrix};\quad B_1 = \begin{bmatrix} -\sqrt{\kappa_2}I \\ 0 \end{bmatrix};\\
&\qquad B_2 = \begin{bmatrix} -\sqrt{\kappa_3}I \\ -\sqrt{\alpha}I \end{bmatrix};\\
&\qquad (\gamma = \kappa_1+\kappa_2+\kappa_3)\\
C_1 &= \begin{bmatrix} \sqrt{\kappa_3}I & 0 \end{bmatrix};\quad D_{12} = I;\\
C_2 &= \begin{bmatrix} \sqrt{\kappa_2}I & 0 \end{bmatrix};\quad D_{20} = 0;\quad D_{21} = I.
\end{aligned}
\tag{8.34}
$$

Here α and β are parameters of the optical amplifier. The signals have Ito matrices $F_u = F_{\tilde{w}} = I + iJ$ and $F_v = \mathrm{diag}(I + iJ, (2N+1)I + iJ)$, and the parameters are chosen to be $\kappa_1 = 2.6$, $\kappa_2 = \kappa_3 = 0.2$, $\alpha = 1$ and $\beta = 0.5$.

With a H^∞ gain $g = 0.1$, the Riccati equations (8.12) and (8.13) have stabilizing solutions satisfying Assumption 8.3.3: $X = Y = 0_{2\times 2}$. Using (8.14), the controller matrices A_K, B_K, C_K are

$$
\begin{aligned}
A_K &= \begin{bmatrix} -1.3894I & -0.4472I \\ -0.2I & -0.25I \end{bmatrix},\ B_K = \begin{bmatrix} -0.4472I \\ 0_{2\times 2} \end{bmatrix},\\
C_K &= \begin{bmatrix} -0.4472I & 0_{2\times 2} \end{bmatrix}.
\end{aligned}
\tag{8.35}
$$

We would like to realize a controller with both classical and quantum degrees of freedom. Suppose further that we want $\xi_c = (\xi_1, \xi_2)^T$ to be quantum and $\xi_q = (\xi_3, \xi_4)$ be classical. For the sake of conformity with the setup of Chapter 7, let

us relabel ξ according $(\xi_1, \xi_2, \xi_3, \xi_4)^T \mapsto (\xi_3, \xi_4, \xi_1, \xi_2)^T = (\xi_c, \xi_q)^T$. Subsequently, we also redefine A_K, B_K, C_K to be:

$$A_K = \begin{bmatrix} -0.25I & -0.2I \\ -0.4472I & -1.3894I \end{bmatrix}, \; B_K = \begin{bmatrix} 0_{2\times 2} \\ -0.4472I \end{bmatrix},$$
$$C_K = \begin{bmatrix} 0_{2\times 2} & -0.4472I \end{bmatrix}. \tag{8.36}$$

Therefore, for our realization we choose $\Theta_K = \mathrm{diag}(0_{2\times 2}, J)$. The realization of the controller is shown in Figure 8.12, which consists of a four-mirror optical cavity, a classical system, and homodyne detection and modulation for interfacing the classical and quantum components. The quantum noises in Figure 8.12 are all canonical. The cavity has coupling coefficients $\kappa_{K1} = \kappa_{K3} = \kappa_{K4} = 0.2$ and $\kappa_{K2} = 2.1788$. The interconnection fields are given by $d\eta_q = \sqrt{1.33}\xi_q dt + dv_{K2}$, and $d\zeta_q = \zeta_c dt + dv_{K4}$, where $\eta_c = (\eta_{c1}, \eta_{c2})^T = (\eta_{q1} - v_{K31}, \eta_{q2} + v_{K32})^T$. For this realization we have the dynamics (8.9) with

$$B_{K1} = \begin{bmatrix} 0.4472I & -1.4761I & 0_{2\times 2} & -0.4472I \\ 0_{2\times 2} & -0.1355I & 0.1355\tilde{I} & 0_{2\times 2} \end{bmatrix},$$

$$B_{K0} = \begin{bmatrix} I & 0_{2\times 2} & 0_{2\times 2} \end{bmatrix},$$

and $v_K = (v_{K1}, v_{K2}, v_{K3}, v_{K4})^T$, where we recall that $\tilde{I} = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}$. To connect this direct realization with the Lemma 8.3.6, consider the following augmentation of the matrices A_K, B_K, C_K in (8.36):

$$\tilde{A_K} = \begin{bmatrix} A_K & 0_{2\times 2} \\ 0_{4\times 2} & 0.4472J \quad 0.25I \end{bmatrix} \quad \tilde{B}_K = \begin{bmatrix} B_K \\ 0_{2\times 2} \end{bmatrix}$$
$$\tilde{C}_K = \begin{bmatrix} C_K & 0_{2\times 2} \end{bmatrix}.$$

Let P be a permutation matrix such that $P\tilde{\Theta}_K P^T = \mathrm{diag}_2(J)$. Then applying

Lemma 8.3.6 to the triple $\{P\tilde{A}_K P^T, P\tilde{B}, \tilde{C}P^T\}$ and choosing

$$\Xi = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0092 & 0 & 0 & 0 & 0.05 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0092 & -0.05 & 0 \\ 0 & 0 & 0 & -0.05 & 0.5947 & 0 \\ 0 & 0.05 & 0 & 0 & 0 & 0.5947 \end{bmatrix}$$

$$\Lambda_{b1} = \begin{bmatrix} 0 & 0.0678i & 0 & -0.0678 & 0.738 & 0.738i \\ 0 & -0.0678i & 0 & -0.0678 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.2236 & 0.2236i \end{bmatrix},$$

gives us the controller realization of Figure 8.12.

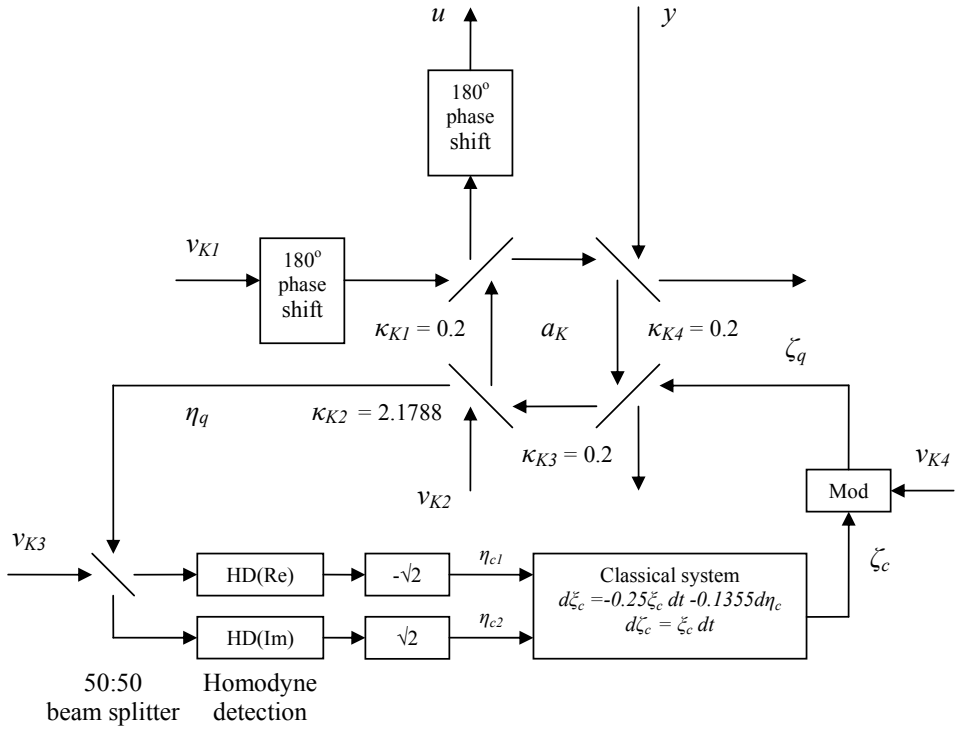

Figure 8.12: Quantum-classical controller ($\Theta_K = \text{diag}(0_{2\times 2}, J)$) for the plant of Figure 8.11.

The above is, of course, not the only possible mixed quantum-classical realization. To obtain a different realization we could have alternatively specified that $(\xi_1, \xi_2)^T$ be classical and $(\xi_3, \xi_4)^T$ as quantum. Let us now do this and set $\xi_c = (\xi_1, \xi_2)^T$ and $\xi_q = (\xi_3, \xi_4)^T$. In this case we are already in the setup of

Chapter 7 and it is not necessary to relabel ξ or redefine A_K, B_K, C_K. In other words, A_K, B_K, C_K are as given in (8.35). This alternative realization is shown in Figure 8.13.

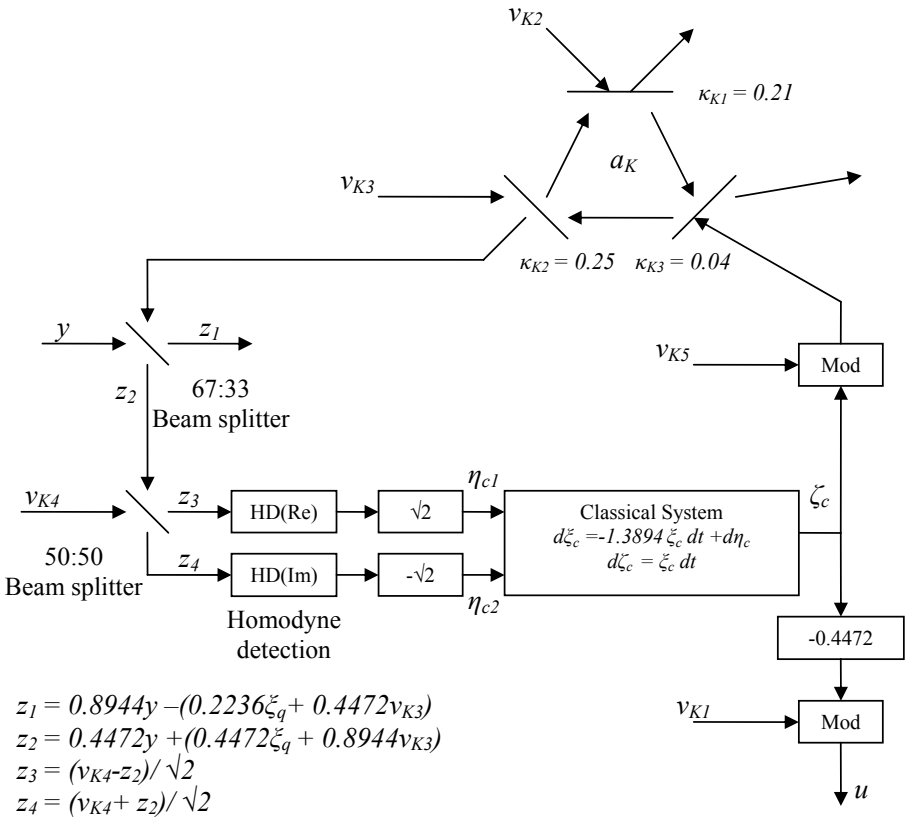

Figure 8.13: Alternative quantum-classical controller ($\Theta_K = \mathrm{diag}(0_{2\times 2}, J)$) for the plant of Figure 8.11.

Note that the control signal u is now the output of the classical part of the controller, modulated by the Wiener noise v_{K1}. As before, we proceed to relate this alternative mixed classical-quantum realization with Lemma 8.3.6. The suitable augmentation of A_K, B_K, C_K are as follows:

$$\begin{aligned}
\tilde{A_K} &= \begin{bmatrix} A_K & 0_{2\times 2} \\ -0.1J \quad 0.2J & 1.3894I \end{bmatrix} \quad \tilde{B}_K = \begin{bmatrix} B_K \\ 0_{2\times 2} \end{bmatrix} \\
\tilde{C}_K &= \begin{bmatrix} C_K & 0_{2\times 2} \end{bmatrix}.
\end{aligned}$$

Then application of Lemma 8.3.6 to $\{P\tilde{A}_K P^T, P\tilde{B}_K, \tilde{C}_K P^T\}$ with

$$\Xi = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.45 & 0 & 0 & 0 & 0.1118 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.45 & -0.1118 & 0 \\ 0 & 0 & 0 & -0.1118 & 0.125 & 0 \\ 0 & 0.1118 & 0 & 0 & 0 & 0.0125 \end{bmatrix}$$

$$\Lambda_{b1} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.2291i & 0.2291i \\ 0 & 0.4472i & 0 & -0.4472 & 0.25 & 0.25i \\ 0 & -0.5i & 0 & -0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.1 & 0.1 \end{bmatrix},$$

results in the controller of Figure 8.13.

8.5 Concluding remarks and future challenges

In this chapter we describe an H^∞ synthesis problem for a class of non-commutative stochastic models due to James and Petersen [32], and then show that the resulting partial description of a controller from the H^∞ synthesis can always be completed, by appropriately adding additional quantum noise channels, such that the complete controller model is physically realizable. It should be emphasized that the novelty of the approach discussed herein is that it facilitates the design of controllers which may have quantum freedom degrees of freedom, such as in [87]. Moreover, the H^∞ theory is applicable to a fairly general class of linear non-commutative plants as well as being quite similar to the classical theory.

A particularly interesting insight obtained from our results is there is complete freedom to specify the type of controller to be realized. Therefore, starting with a partial description of a controller, one may at will choose to realize either a fully quantum controller (with no classical components), a classical controller (with no quantum components), or even a more elaborate hybrid type of controller with mixed classical and quantum components. All these different types of realization will achieve the same H^∞ performance specification. In a number of illustrative examples from quantum optics, we demonstrate the synthesis and physical realization of quantum, classical, and quantum-classical controllers.

The initial results and insights of this work opens several avenues for future research in quantum linear stochastic systems. We shall now list and discuss some

interesting interrelated topics which can be themes for further investigations:

1. Fortuitously, the resulting system matrices in the synthesis examples have a special structure consisting of blocks of scalar multiples of 2×2 identity matrices, thus enabling direct (or "manual") realization the controller. For a number of these direct realizations we also explicitly show how they correspond to one of the infinitely many different realizations that can be obtained from Theorem 8.3.5. However, there is no reason to believe that we will in general obtain solutions with readily exploitable structures nor would direct realization necessarily be sensible for a controller with many degrees of freedom. The general approach would be to apply Theorem 8.3.5 and Lemma 8.3.6 to obtain a pair of Hamiltonian and coupling matrices (R, Λ) which completely specifies a particular realization, followed by a physical construction of the controller. At present, an impediment to this approach is, to the best of the author's knowledge, a lacuna of results pertaining to systematic engineering of linear quantum optical systems with arbitrarily specified (R, Λ) from a bin of quantum optical components. There is an interest in the quantum information science community in implementation of quadratic Hamiltonians but only of specific ones which will create entangled Gaussian states useful for quantum cryptography and computing [112, 113, 114]. In [115], it has been shown that arbitrary quadratic Hamiltonians can be implemented via applying appropriate sequences of fast elementary operations and is investigated further in, e.g., [112, 113]. Despite this, it may still prove advantageous to be able to engineer the Hamiltonian directly and in some situations this may be preferable. On the other hand, engineering of arbitrary linear coupling seems to have attracted even less attention, if any. With the advent of quantum control and the increasing involvement of engineers, these implementation questions may in the future gain more prominence. In any case, they are important questions and will be even more relevant if we are to develop other synthesis methods in which we do not have the luxury of adding in quantum noise without degrading the closed loop performance. It is hoped that this will eventually lead to a constructive theory for the analysis and synthesis of quantum optical networks which parallels that of electrical networks.

2. Due to the presence of the free parameter Ξ in Lemma 8.3.6, we know that there can be infinitely many different physical realizations for any particular

partial model. As such it would also be of interest to consider whether it is possible to find Ξ which returns a realization which is in some sense "optimal". For example, it may be that one wishes to find a realization with a minimal number of noise channels, that is, a realization with N_{v_K} which is smallest among all possible realizations. Alternatively, we may wish to bound the complexity of the controller and look for a realization which can be built with a minimum number of components.

3. Here we have only considered a quantum generalization of classical H^∞ synthesis. However, there are other controller synthesis paradigms such as the LQG synthesis and the risk-sensitive synthesis. Thus, it is of interest to investigate whether these other paradigms may also be extended to the quantum linear stochastic systems setting.

Bibliography

[1] E. Wong and B. Hajek, *Stochastic Processes in Engineering Systems.* Springer-Verlag, New York, 1985.

[2] F. M. Callier and J. Winkin, "The spectral factorization problem for SISO distributed systems," in *Modelling, Robustness and Sensitivity Reduction in Control Theory*, ser. NATO ASI, R. F. Curtain, Ed. Springer-Verlag, 1987, vol. F34, pp. 463–489.

[3] ——, "Spectral factorization and LQ-optimal regulation for multivariable distributed systems," *Internat. J. Control*, vol. 52, pp. 55–75, 1990.

[4] O. J. Staffans, "Quadratic optimal control of stable systems through spectral factorization," *Math. Control Signals Systems*, vol. 8, pp. 167–197, 1995.

[5] C. I. Byrnes, S. V. Gusev, and A. Lindquist, "A convex optimization approach to the rational covariance extension problem," *SIAM J. Control Optim.*, vol. 37, no. 1, pp. 211–229, 1998.

[6] ——, "From finite covariance windows to modeling filters: a convex optimization approach," *SIAM Rev.*, vol. 43, no. 4, pp. 645–675, 2001.

[7] M. Pourahmadi, *Foundations of time series analysis and prediction theory*, ser. Wiley series in Probability and statistics. John Wiley & Sons, Inc., 2001.

[8] P. Dewilde, A. C. Viera, and T. Kailath, "On a generalized Szegö-Levinson realization algortihm for optimal linear predictors based on a network synthesis approach," *IEEE Trans. Circuits Syst.*, vol. 25, no. 9, pp. 663–675, Sept. 1978.

[9] T. T. Georgiou, "Realization of power spectra from partial covariance sequences," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 42, no. 4, pp. 1049–1070, Apr. 1987.

[10] D. C. Youla and N. N. Kazanjian, "Bauer-type factorization of positive matrices and the theory of matrix polynomials orthogonal on the unit circle," *IEEE Trans. Circuits Syst.*, vol. 25, no. 2, pp. 57–69, Feb. 1978.

[11] P. Delsarte, Y. V. Genin, and Y. G. Kamp, "Orthogonal polynomial matrices on the unit circle," *IEEE Trans. Circuits Syst.*, vol. 25, no. 3, pp. 149–160, Mar. 1978.

[12] T. T. Georgiou, "Partial realization of covariance sequences," Ph.D. dissertation, Dept. of Elec. Eng., Center for Mathematical Systems Theory, University of Florida, 1983.

[13] J. Mari, A. Dahlèn, and A. Lindquist, "A covariance extension approach to identification of time series," *Automatica J. IFAC*, vol. 36, pp. 379–398, 2000.

[14] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. Automat. Contr.*, vol. 46, no. 5, pp. 822–839, May 2001.

[15] S. Barclay, "Continuity of the spectral factorization mapping," *J. London Math. Soc. (2)*, vol. 70, pp. 763–779, Dec. 2004.

[16] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev, "A complete parametrization of all positive rational extensions of a covariance sequence," *IEEE Trans. Automat. Contr.*, vol. 40, no. 11, pp. 1841–1857, Nov. 1995.

[17] H. I. Nurdin and A. Bagchi, "On the solutions of the rational covariance extension problem corresponding to pseudopolynomials having boundary zeros," *IEEE Trans. Automat. Contr.*, vol. 51, no. 2, pp. 350–355, 2006.

[18] H. I. Nurdin, "New results on the rational covariance extension problem with degree constraint," *Systems Control Lett.*, vol. 55, no. 7, pp. 530–537, July 2006. [Online]. Available: http://dx.doi.org/10.1016/j.sysconle.2005.11.004

[19] B. D. O. Anderson, "Continuity of the spectral factorization operation," *Mat. Aplic. Comp.*, vol. 4, no. 2, pp. 139–156, 1985.

[20] P. Caines and M. Baykal-Gürsoy, "On the L^∞ consistency of L^2 estimators," *Systems Control Lett.*, vol. 12, pp. 71–76, 1989.

[21] M. C. Roggemann and B. Welsh, *Imaging Through Turbulence*. Boca Raton: CRC Press LLC, 1996.

[22] J. M. Conan, G. Rosset, and P. Y. Madec, "Wave-front temporal spectra in high resolution imaging through turbulence," *J. Opt. Soc. Am. A*, vol. 12, no. 7, pp. 1559–1570, July 1995.

[23] D. Buscher, J. T. Armstrong, C. A. Hummel, A. Quirrenbach, D. Mozurkewich, K. J. Johnston, C. S. Denison, M. M. Colavita, and M. Shao, "Interferometric seeing measurements on Mt. Wilson: Power spectra and outer scales," *Applied Optics*, vol. 34, no. 6, pp. 1081–1096, February 1995.

[24] D. R. McGaughey and G. J. M. Aitken, "Temporal analysis of stellar wave-front-tilt data," *J. Opt. Soc. Am. A*, vol. 14, no. 8, pp. 1967–1974, August 1997.

[25] R. Rao, S. Wang, X. Liu, and Z. Gong, "Turbulence spectrum effect on wave temporal-frequency spectra for light propagating through the atmosphere," *J. Opt. Soc. Am. A*, vol. 16, no. 11, pp. 2755–2762, November 1999.

[26] B. Le Roux, J. M. Conan, C. Kulcsár, H. F. Raynaud, L. M. Mugnier, and T. Fusco, "Optimal control law for classical and multiconjugate adaptive optics," *J. Opt. Soc. Am. A*, vol. 21, no. 7, pp. 1261–1276, July 2004.

[27] D. P. Looze, "Minimum variance control structure for adaptive optics system," *J. Opt. Soc. Am. A*, vol. 23, no. 3, pp. 603–612, March 2006.

[28] U. L. Ly and Y. K. Chan, "Time- domain computation of aircraft gust covariance matrices," in *Atmospheric Flight Mechanics Conference, Danvers, Mass., August 11-13, 1980, Collection of Technical Papers.* New York: AIAA, Inc., 1980, pp. 482–491.

[29] S. Gage, "Creating a unified graphical wind turbulence model from multiple specifications," in *AIAA Modeling and Simulation Technologies Conference*

and Exhibit, Austin, Texas, August 11-14, 2003. AIAA, Inc., August 2003, paper no. AIAA-2003-5529.

[30] A. V. Balakrishnan, *Introduction to Random Processes in Engineering*. John Wiley & Sons, Inc, 1995.

[31] B. D. O. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis: A Modern Systems Theory Approach*, ser. Networks Series. Prentice-Hall, Inc., 1973.

[32] M. R. James and I. R. Petersen, "H^∞ control of linear quantum systems," in *Proceedings of the 45th IEEE Conference on Decision and Control (San Diego, CA, December 13-15, 2006)*, 2006, pp. 3126–3131.

[33] U. Grenander and G. Szegö, *Toeplitz Forms and Their Applications*, ser. California monographs in mathematical sciences. Berkeley and Los Angeles: University of California Press, 1958.

[34] P. J. Brockwell and R. A. Davis, *Time series: Theory and methods*, 2nd ed., ser. Springer series in statistics. Springer, 1991.

[35] P. Dewilde and H. Dym, "Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences," *IEEE Trans. Inform. Theory*, vol. 27, no. 4, pp. 446–461, July 1981.

[36] ——, "Lossless chain scattering matrices and optimum linear prediction: The vector case," *Circuit Theory Appl.*, vol. 9, pp. 135–175, 1981.

[37] ——, "Lossless inverse scattering, digital filters, and estimation theory," *IEEE Trans. Inform. Theory*, vol. 30, no. 4, pp. 644–662, July 1984.

[38] P. Delsarte and Y. Genin, "On a generalization of the Szegö-Levinson recurrence and its application in lossless inverse scattering," *IEEE Trans. Inform. Theory*, vol. 38, no. 1, pp. 104–110, Jan. 1992.

[39] N. Wiener and P. Masani, "The predicition theory of multivariate stochastic processes: I. The regularity condition," *Acta Math.*, vol. 98, pp. 111–150, 1957.

[40] ——, "The prediction theory of multivariate stochastic processes: II. The linear predictor," *Acta Math.*, vol. 99, pp. 93–137, 1958.

[41] P. Masani, "The prediction theory of multivariate stochastic processes: III. Unbounded spectral densities," *Acta Math.*, vol. 104, pp. 141–162, 1960.

[42] H. Helson and D. Lowdenslager, "Prediction Theory and Fourier Series in Several Variables," *Acta Math.*, vol. 99, pp. 165–202, 1958.

[43] L. Y. Geronimus, *Orthogonal Polynomials*. New York: Consultants Bureau, 1961, authorized translation from the Russian.

[44] N. Wiener, "On the factorization of matrices," *Comment. Math. Helv.*, vol. 29, pp. 97–111, 1955.

[45] P. Whittle, "On the fitting of multivariate autoregressions, and the approximate canonical factorization of a spectral density matrix," *Biometrika*, vol. 50, no. 1/2, pp. 129–134, June 1963.

[46] N. I. Akhiezer, *The classical moment problem and some related questions in analysis*. Edinburgh and London: Oliver & Boyd, 1965.

[47] P. Dewilde, "How to approximate a spectrum recursively using ARMA models," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP '82)*. IEEE, May 1982, pp. 1736–1739.

[48] T. T. Georgiou, "The interpolation problem with degree constraint," *IEEE Trans. Automat. Contr.*, vol. 44, no. 3, pp. 631–635, Mar. 1999.

[49] A. Blomqvist, G. Fanizza, and R. Nagamune, "Computation of bounded degree Nevanlinna-Pick interpolants by solving nonlinear equations," in *Proceedings of the 42nd IEEE Conference on Decision and Control*, 2003, pp. 4511–4516.

[50] P. Enqvist, "A homotopy approach to rational covariance extension with degree constraint," *Int. J. Appl. Math. Comput. Sci.*, vol. 11, no. 5, pp. 1173–1201, 2001.

[51] C. I. Byrnes, T. T. Georgiou, A. Lindquist, and A. Megretski, "Generalized interpolation in $\mathcal{H}^\infty$ with a complexity constraint," *Trans. Amer. Math. Soc.*, vol. 358, pp. 965–987, 2006.

[52] H. I. Nurdin and A. Bagchi, "On the solutions of the rational covariance extension problem corresponding to pseudopolynomials having boundary zeros," in *Proceedings of the 43rd IEEE Conference on Decision and Control, Paradise Island, The Bahamas*, 2004, pp. 5386–5391.

[53] J. L. Walsh, *Interpolation and Approximation by Rational Functions in the Complex Domain*, 2nd ed. Ann Arbor, MI: Edwards, 1956.

[54] P. Billingsley, *Probability and Measure*, 2nd ed., ser. Wiley Series in Probability and Mathematical Statistics. Wiley, 1986.

[55] A. Blomqvist, A. Lindquist, and R. Nagamune, "Matrix-valued Nevanlinna-Pick interpolation with complexity constraint: An optimization approach," *IEEE Trans. Automat. Contr.*, vol. 48, no. 12, pp. 2172–2190, Dec. 2003.

[56] G. Heining, "Inversion of Toeplitz-plus-Hankel matrices with arbitrary rank profile," in *Contemporary Mathematics*. American Mathematical Society, 2003, vol. 323, pp. 75–88.

[57] G. Heinig and K. Rost, "New fast algorithms for Toeplitz-plus-Hankel matrices," *SIAM J. Matrix Anal. Appl.*, vol. 25, no. 3, pp. 842–857, 2004.

[58] A. Blomqvist and R. Nagamune, "Optimization-based computation of analytic interpolants of bounded complexity," *Syst. Control Lett.*, vol. 54, no. 855-864, 2005.

[59] R. Nagamune, "A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint," *IEEE Trans. Automat. Contr.*, vol. 48, no. 1, pp. 113–117, Jan 2003.

[60] A. Blomqvist and R. Nagamune, "An extension of a Nevanlinna-Pick interpolation solver to cases including derivative constraints," in *Proceedings of the 41st IEEE Conference in Decision and Control, Las Vegas*, Dec 2002.

[61] A. Blomqvist, G. Fanizza, and R. Nagamune, "Computation of bounded degree Nevanlinna-Pick interpolants by solving nonlinear equations," Rep. no. 17, 2003, Mittag Leffler Preprint Series, Spring 2002/2003.

[62] C. I. Byrnes and A. Lindquist, "The uncertain generalized moment problem with complexity constraint," in *New Trends in Nonlinear Dynamics and*

Control, W. Kang, M. Xiao, and C. Borges, Eds. Springer-Verlag, 2003, pp. 267–278.

[63] ——, "The generalized moment problem with complexity constraint," *Integral Equations and Operator Theory*, vol. 56, no. 2, 2006.

[64] A. H. Sayed and T. Kailath, "A survey of spectral factorization methods," *Numer. Linear Algebra Appl.*, vol. 08, pp. 467–496, 2001.

[65] P. L. Duren, *Theory of H^p spaces*. Mineola, New York: Dover Publications, Inc., 2000.

[66] H. I. Nurdin, "A new approach to spectral factorization of a class of matrix-valued spectral densities," in *Proceedings of the Joint 44th IEEE Conference on Decision and Control (CDC) and European Control Conference (ECC)*, Dec. 2005, pp. 5929–5934.

[67] C. I. Byrnes and A. Lindquist, "On the duality between filtering and Nevanlinna-Pick interpolation," *SIAM J. Control Optim.*, vol. 39, no. 3, pp. 757–775, 2000.

[68] C. I. Byrnes, H. J. Landau, and A. Lindquist, "On the well-posedness of the rational covariance extension problem," in *Current and Future Directions in Applied Mathematics*, M. Alber, B. Hu, and J. Rosenthal, Eds. Boston: Birkhäuser, 1997, pp. 81–108.

[69] T. T. Georgiou, "Spectral estimation via selective harmonic amplification," *IEEE Trans. Automat. Contr.*, vol. 46, no. 1, pp. 29–42, Jan 2001.

[70] C. I. Byrnes, T. T. Georgiou, and A. Lindquist, "A new approach to spectral estimation: A tunable high-resolution spectral estimator," *IEEE Trans. Signal Processing*, vol. 49, pp. 3189–3205, 2000.

[71] R. Nagamune, "Closed-loop shaping based on Nevanlinna-Pick interpolation with a degree bound," *IEEE Trans. Automat. Contr.*, vol. 49, no. 2, pp. 300–305, Feb 2004.

[72] ——, "A robust solver using a continuation method for Nevanlinna-Pick interpolation with degree constraint," in *Proceedings of the 40th IEEE Conference on Decision and Control, Orlando, FL*, December 2001, pp. 1119–1124.

[73] H. I. Nurdin and J. B. Moore, "Computation of degree constrained rational interpolants with non-strictly positive parametrizing functions via homotopy continuation," in *Proceedings of the 45th IEEE Conference on Decision and Control (San Diego, CA, Dec. 13-15, 2006)*, 2006, pp. 565–570.

[74] M. Fu and K. Mahata, "Generalizations of the Nevanlinna-Pick interpolation problem," in *Proceedings of the Joint 44th IEEE Conference on Decision and Control (CDC) and European Control Conference (ECC)*, Dec. 2005, pp. 4299–4304.

[75] M. Pavon and A. Ferrante, "On the Georgiou-Lindquist approach to constrained Kullback-Leibler approximation of spectral densities," *IEEE Trans. Automat. Contr.*, vol. 51, pp. 639–644, Feb. 2006.

[76] L. Ephremidze, G. Janashia, and E. Lagvilava, "On approximate factorization of positive definite matrix functions," *Uspekhi Mat. Nauk*, vol. 54, no. 1, pp. 161–162, 1999, in Russian, translation in *Russian Math. Surveys*, (54) 6, pp. 1246-1247.

[77] H. I. Nurdin, "Spectral factorization of a class of matrix-valued spectral densities," *SIAM J. Control Optim.*, vol. 48, no. 5, pp. 1801–1821, 2006.

[78] T. W. Körner, *Fourier Analysis*. Cambridge University Press, 1988.

[79] R. T. Rockafellar, *Conjugate Duality and Optimization*, ser. Regional Conference Series in Applied Mathematics. SIAM, 1974, no. 16.

[80] H. F. Davis, *Fourier series and orthogonal functions*. Boston: Allyn and Bacon, Inc., 1963.

[81] A. V. Balakrishnan, "A general formulation and solution to the A-O control problem," 2000, personal communication.

[82] V. Belavkin, "On the theory of controlling observable quantum systems," *Automation and Remote Control*, vol. 44, no. 2, pp. 178–188, 1983.

[83] H. Wiseman, "Quantum theory of continuous feedback," *Phys. Rev. A*, vol. 49, no. 3, pp. 2133–2150, 1994.

[84] A. Doherty and K. Jacobs, "Feedback-control of quantum systems using continuous state-estimation," *Phys. Rev. A*, vol. 60, p. 2700, 1999, quant-ph/9812004.

[85] M. Armen, K. Au, J. Stockton, A. Doherty, and H. Mabuchi, "Adaptive homodyne measurement of optical phase," *Phys. Rev. A*, vol. 89, no. 13, p. 133602, 2002.

[86] M. Yanagisawa and H. Kimura, "Transfer function approach to quantum control-part I: Dynamics of quantum feedback systems," *IEEE Trans. Automatic Control*, vol. 48, no. 12, pp. 2107–2120, 2003.

[87] ——, "Transfer function approach to quantum control-part II: Control concepts and applications," *IEEE Trans. Automatic Control*, vol. 48, no. 12, pp. 2121–2132, 2003.

[88] J. Geremia, J. Stockton, and H. Mabuchi, "Real-time quantum feedback control of atomic spin-squeezing," *Science*, vol. 304, pp. 270–273, April 2004.

[89] M. Nielsen and I. Chuang, *Quantum Computation and Quantum Information*. Cambridge: Cambridge University Press, 2000.

[90] E. Knill, R. Laflamme, and G. Milburn, "A scheme for efficient quantum computation with linear optics," *Nature*, vol. 409, pp. 46–52, 2001.

[91] R. L. Hudson and K. R. Parthasarathy, "Quantum Ito's formula and stochastic evolution," *Commun. Math. Phys.*, vol. 93, pp. 301–323, 1984.

[92] K. Parthasarathy, *An Introduction to Quantum Stochastic Calculus*. Berlin: Birkhauser, 1992.

[93] C. Gardiner and P. Zoller, *Quantum Noise: A Handbook of Markovian and Non-Markovian Quantum Stochastic Methods with Applications to Quantum Optics*, 2nd ed., ser. Springer Series in Synergetics. Springer, 2000.

[94] L. Bouten, R. van Handel, and M. R. James, "An introduction to quantum filtering," 2006, arxiv.org/math.OC/0601741.

[95] I. Petersen, B. Anderson, and E. Jonckheere, "A first principles solution to the non-singular H^∞ control problem," *International Journal of Robust and Nonlinear Control*, vol. 1, no. 3, pp. 171–185, 1991.

[96] E. Merzbacher, *Quantum Mechanics*, 3rd ed. New York: Wiley, 1998.

[97] D. Petz, *An Invitation to the Algebra of Canonical Commutation Relations*, ser. Series A: Mathematical Physics of *Leuven Notes in Mathematical and Theoretical Physics*. Leuven, Belgium: Leuven University Press, 1989, vol. 2.

[98] C. D'Helon and M. James, "Stability, gain, and robustness in quantum feedback networks," *Phys. Rev. A*, vol. 73, pp. 0 532 803–1–0 532 803–1, 2006, arxiv.org/quant-ph/0511140.

[99] A. Holevo, *Statistical Structure of Quantum Theory*. Berlin: Springer, 2001.

[100] M. R. James, H. I. Nurdin, and I. R. Petersen, "H^∞ control of linear quantum stochastic systems," 2006, submitted for publication. [Online]. Available: arxiv.org/quant-ph/0703150

[101] V. Belavkin, "Continuous non-demolition observation, quantum filtering and optimal estimation," in *Quantum Aspects of Optical Communication*, ser. Lecture notes in Physics. Berlin: Springer, 1991, vol. 45, pp. 131–145.

[102] S. C. Edwards and V. P. Belavkin, "Optimal quantum filtering and quantum feedback control," August 2005. [Online]. Available: http://arxiv.org/pdf/quant-ph/0506018

[103] E. Davies, *Quantum Theory of Open Systems*. New York: Academic Press, 1976.

[104] J. Willems, "Dissipative dynamical systems - Part I: General theory," *Archive of Rational Mechanics and Analysis*, vol. 45, pp. 321–351, 1972.

[105] A. van der Schaft, *L_2-Gain and Passivity Techniques in Nonlinear Control*. New York: Springer Verlag, 1996.

[106] P. Dupuis, M. James, and I. Petersen, "Robust properties of risk-sensitive control," *Math. Control, Systems and Signals*, vol. 13, pp. 318–332, 2000.

[107] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan, *Linear Matrix Inequalities in System and Control Theory*. Philadelphia, Pennsylvania: SIAM, 1994.

[108] K. Zhou and P. Khargonekar, "An algebraic Riccati equation approach to H_∞ optimization," *Systems and Control Letters*, vol. 11, pp. 85–91, 1988.

[109] M. Vidyasagar, *Nonlinear Systems Analysis*, 2nd ed. Englewood Cliffs: Prentice-Hall, 1993.

[110] J. Doyle, K. Glover, P. Khargonekar, and B. Francis, "State-space solutions to the standard H_2 and H_∞ control problems," *IEEE Transactions on Automatic Control*, vol. 34, no. 8, pp. 831–847, 1989.

[111] H. Bachor and T. Ralph, *A Guide to Experiments in Quantum Optics*, 2nd ed. Weinheim, Germany: Wiley-VCH, 2004.

[112] B. Kraus, K. Hammerer, G. Giedke, and J. I. Cirac, "Entanglement generation and Hamiltonian simulation in continuous-variable systems," *Phys. Rev. A*, vol. 67, pp. 042 314–1–042 314–14, 2003.

[113] J. Fiurášek, "Unitary gate synthesis for continuous variable systems," *Phys. Rev. A*, vol. 68, pp. 022 304–1–022 304–8, 2003.

[114] D. Dell'Anno, S. De Siena, and F. Illuminati, "Multiphoton quantum optics and quantum state engineering," *Phys. Rep*, vol. 428, pp. 53–168, 2006, arxiv.org/quant-ph/0701050.

[115] S. Lloyd and S. L. Braunstein, "Quantum computation over continuous variables," *Phys. Rev. Lett*, vol. 82, no. 8, pp. 1785–1787, February 1999.