Inverse Problems in Analytic Interpolation for Robust Control and Spectral Estimation

JOHAN KARLSSON

Doctoral Thesis
Stockholm, Sweden 2008

TRITA MAT 08/OS/09
ISSN 1401-2294
ISRN KTH/OPT SYST/DA-08/09-SE
ISBN 978-91-7415-125-1

Optimization and Systems Theory
Department of Mathematics
Royal Institute of Technology
SE-100 44 Stockholm, SWEDEN

Akademisk avhandling som med tillstånd av Kungl Tekniska Högskolan framlägges till offentlig granskning för avläggande av teknologie doktorsexamen, fredagen den 31 oktober 2008 klockan 13.00 i rum F3, Lindstedtsvägen 26, Kungl Tekniska Högskolan, Stockholm.

Print: Kista Snabbtryck AB

Although this may seem a paradox,
all exact science is dominated by the idea of approximation.

- Bertrand Russell

Till min familj

Abstract

This thesis is divided into two parts. The first part deals with the Nevanlinna-Pick interpolation problem, a problem which occurs naturally in several applications such as robust control, signal processing and circuit theory. We consider the problem of shaping and approximating solutions to the Nevanlinna-Pick problem in a systematic way. In the second part, we study distance measures between power spectra for spectral estimation. We postulate a situation where we want to quantify robustness based on a finite set of covariances, and this leads naturally to considering the weak*-topology. Several weak*-continuous metrics are proposed and studied in this context.

In the first paper we consider the correspondence between weighted entropy functionals and minimizing interpolants in order to find appropriate interpolants for, e.g., control synthesis. There are two basic issues that we address: we first characterize admissible shapes of minimizers by studying the corresponding inverse problem, and then we develop effective ways of shaping minimizers via suitable choices of weights. These results are used in order to systematize feedback control synthesis to obtain frequency dependent robustness bounds with a constraint on the controller degree.

The second paper studies contractive interpolants obtained as minimizers of a weighted entropy functional and analyzes the role of weights and interpolation conditions as design parameters for shaping the interpolants. We first show that, if, for a sequence of interpolants, the values of the corresponding entropy gains converge to the optimum, then the interpolants converge in H_2, but not necessarily in H_∞. This result is then used to describe the asymptotic behaviour of the interpolant as an interpolation point approaches the boundary of the domain of analyticity.

A quite comprehensive theory of analytic interpolation with degree constraint, dealing with rational analytic interpolants with an a priori bound, has been developed in recent years. In the third paper, we consider the limit case when this bound is removed, and only stable interpolants with a prescribed maximum degree are sought. This leads to weighted H_2 minimization, where the interpolants are parameterized by the weights. The inverse problem of determining the weight given a desired interpolant profile is considered, and a rational approximation procedure based on the theory is proposed. This provides a tool for tuning the solution for attaining design specifications.

The purpose of the fourth paper is to study the topology and develop metrics that allow for localization of power spectra, based on second-order statistics. We show that the appropriate topology is the weak*-topology and give several examples on how to construct such metrics. This allows us to quantify uncertainty of spectra in a natural way and to calculate a priori bounds on spectral uncertainty, based on second-order statistics. Finally, we study identification of spectral densities and relate this to the trade-off between resolution and variance of spectral estimates.

In the fifth paper, we present an axiomatic framework for seeking distances between power spectra. The axioms require that the sought metric respects the effects of additive and multiplicative noise in reducing our ability to discriminate spectra. They also require continuity of statistical quantities with respect to perturbations measured in the metric. We then present a particular metric which abides by these requirements. The metric is based on the Monge-Kantorovich transportation problem and is contrasted to an earlier Riemannian metric based on the minimum-variance prediction geometry of the underlying time-series. It is also being compared with the more traditional Itakura-Saito distance measure, as well as the aforementioned prediction metric, on two representative examples.

Keywords: Nevanlinna-Pick Interpolation, Approximation, Model reduction, Robust Control, Gap-robustness, Sensitivity Shaping, Entropy functional, Spectral Estimation, Weak*-topology, Monge-Kantorovic Transportation Problem.

Acknowledgments

Anders Lindquist, my deepest thanks for sharing your vast mathematical knowledge, extensive experience, and insights from the academics. Your encouragement and support has been invaluable. I am also grateful that you with your dedication and world-wide contact network has created a great working environment at the department.

Tryphon Georgiou, I am sincerely thankful to you. Your enthusiasm, intellect, and constant curiosity has served as fuel injections during my visits in Minneapolis. Your clear thinking and open-mindedness has inspired me and taught me that every detour is worth exploring if you can learn something from it.

I would like to thank my other coauthors whom I had the pleasure to work with. Shahrouz Takyar, for your enthusiasm and generosity. A special thanks for taking care of me during my visits in Minneapolis. I will not forget the Persian parties, our vendettas in the squash court, or the dinners at Applebee's. Per Enqvist, for your insightful and systematic way of doing research. I have enjoyed your company a lot both in Sweden and on our trips all over the world. Thank you also for all the useful feedback on the introduction of this thesis. Ryozo Nagamune, for your devotedness to science and careful and thorough approach to writing and problem solving. Vanna Fanizza, for your sharp eye for identifying useful results and good intuition for mathematics, and I wish you the best of luck with your financial career in Zürich.

I have enjoyed being a PhD-student at the division of Optimization and Systems Theory, and the faculty members Anders Forsgren, Per-Olof Gutman, Xiaoming Hu, Marie Lundin, Krister Svanberg, and Claes Trygger deserve thanks for creating such a enjoyable and stimulating working environment. In particular, thank you Ulf Jönsson, for your careful and unselfish work at the department, and that you despite a heavy work-load always had time for discussions and for giving advice.

I would like to thank my fellow PhD-students whose company I have enjoyed a lot. In particular, David Anisi, for your genuine personality who is always fun to be around. Enrico Avventi, for all the interesting lunch discussions and for your hospitality in Padova. Anders Blomqvist, I enjoyed sharing office with you. I would also like to thank you for always taking the time to explain both the theory and how to solve practical things when I was new at the department. Tove Gustavi, for your enthusiasm and good sportsmanship. No matter if it is squash, skiing or running, you are always up for a good game. Anders Möller, for your positive energy and good sense of humor. Mats Werme, for your generous way of life who always looks for the best of everyone. Thanks Stefan Almér, Fredrik

Carlsson, Mikael Fallgren, and Maja Karasalo, for friendship and the good times we have had during the last five years.

I would also like to thank my other friends and colleagues. In particular, my friends that go way back, which are now organized in måndagsklubben. My friends from bridge, whom I have enjoyed lots of “brickor” with. My friends from salsa and climbing. You all mean very much to me. Thanks Anna, for being such a wonderful person, for constantly challenging me, and for teaching me that it is actually fun to put the head down into the water. Finally, I would like to thank my family, Kjell, Lisbeth, Anna, Susanne, Micke, Rasmus and Lucas, for your endless support and constant kindness. In particular, Mamma och Pappa, thanks also for reminding me of what really matters and for always being there to 100 percent!

Stockholm, October 2008

Johan Karlsson

Contents

Introduction

Mathematical modelling and approximation theory has been at the core of applied mathematics for several centuries. Classical examples of this are Gauss' least square approximation where noisy measurements can be used for approximating a trajectory and Newton's laws which gives a model for movements of masses under gravitational forces. The essence is to obtain representations which provide a simple but yet meaningful description of the real world.

This thesis is divided into two parts. In the first part, we study approximation of systems while preserving specific properties of the system, such as stability, contractiveness, and interpolation values. The underlying mathematical problem appears in many applications, from control synthesis to model reduction of systems. In the second part of the thesis, we study distances and identification of stochastic processes. In particular we explore a topology which allows for robust identification and quantitative descriptions of uncertainty sets.

In this introduction, we describe some of the definitions and concepts that are used in the thesis. First, we consider linear systems, Hardy spaces, and how these can be used in robust control. Then, a brief overview of stochastic processes and spectral estimation follows, and some connections with speech processing are illustrated. Finally, the main contributions of this thesis are presented and the five papers are summarized.

1.1 Linear systems

A class of systems widely used in the literature is the class of linear systems. A discrete-time linear time-invariant (LTI) system h can be represented by an input-output mapping, $h : \mathcal{U} \to \mathcal{Y}$ which takes an input sequence $\{u_k\}_{k\in\mathbb{Z}} \in \mathcal{U}$ to an output sequence $\{y_k\}_{k\in\mathbb{Z}} \in \mathcal{Y}$,

$$y_k = \sum_{\ell\in\mathbb{Z}} h_\ell u_{k-\ell},$$

where $\mathbb{Z}$ denotes the set of integers, and $\{h_k\}_{k\in\mathbb{Z}}$ is specified by h. The input sequence $\{u_k\}_{k\in\mathbb{Z}}$, the output sequence $\{y_k\}_{k\in\mathbb{Z}}$, and the input-output mapping induced by h can be represented by the formal power series

$$u(z) = \sum_k u_k z^k, \quad y(z) = \sum_k y_k z^k, \quad h(z) = \sum_k h_k z^k$$

in which case

$$y(z) = h(z)u(z).$$

Here z represents a unit delay of the signal (the shift-operator). The function $h(z)$ is called the transfer function of the system h, and is represented by a block diagram

$$\text{input} \xrightarrow{u_k} \boxed{h(z)} \xrightarrow{y_k} \text{output.}$$

In this thesis we mainly deal with single-input single-output (SISO) systems, in which case $\{u_k\}_{k\in\mathbb{Z}} \in \mathcal{U}$, $\{y_k\}_{k\in\mathbb{Z}} \in \mathcal{Y}$, and $\{h_k\}_{k\in\mathbb{Z}}$ are sequences of complex numbers. A common requirement on physical systems is causality, i.e., that an output y_k only depends on input u_m for $m \leq k$, and hence $h_k = 0$ for $k < 0$.

A fundamental problem in system identification and model reduction is to find approximate models of physical systems that can be implemented in an efficient way. Due to memory constraints we require the system to be rational, i.e., representable in the form

$$y_k + a_1 y_{k-1} + \cdots + a_n y_{k-n} = \sigma_0 u_k + \sigma_1 u_{k-1} + \cdots + \sigma_m u_{k-m}$$

in which case the transfer function is given by

$$h(z) = \frac{\sigma_0 + \sigma_1 z + \cdots + \sigma_m z^m}{1 + a_1 z + \cdots + a_n z^n} = \frac{\sigma(z)}{a(z)}.$$

The McMillan degree (or simply degree) of a system representable in this form, with σ and a coprime, is $\max(n, m)$. The part of the system specified by a is called the Auto-Regressive (AR) part and the part specified by σ the Moving-Average (MA) part. Therefore we will call a system of the form

$$h(z) = \sigma(z), \quad h(z) = \frac{\sigma_0}{a(z)}, \quad h(z) = \frac{\sigma(z)}{a(z)}$$

a MA-system, an AR-system, and an ARMA-system, respectively. One of the main topics of this thesis is to find ARMA-models of low degree which approximate a given system in a suitable sense.

1.1.1 Norms and Hardy spaces

We consider Hardy spaces H_p which are Banach spaces of analytic functions in the unit disc $\mathbb{D} = \{z : |z| < 1\}$. For $1 \leq p < \infty$, denote by H_p the space of analytic functions f in $\mathbb{D}$ for which the norm

$$\|f\|_{H_p} = \left(\sup_{0\leq r<1} \frac{1}{2\pi} \int_{-\pi}^{\pi} |f(re^{i\theta})|^p d\theta \right)^{\frac{1}{p}}$$

is finite. The space H_∞ is the set of bounded analytic functions in $\mathbb{D}$ with norm

$$\|f\|_{H_p} = \sup_{z\in\mathbb{D}} |f(z)|.$$

Clearly $H_\infty \subset H_p \subset H_q \subset H_1$ when $1 < q < p < \infty$.

Let the input space $\mathcal{U}$ and the output space $\mathcal{Y}$ of a system h be the $\ell_2(\mathbb{Z})$ space, i.e., both contains sequences $\{x_k\}_{k\in\mathbb{Z}}$ for which the norm

$$\|\{x_k\}_{k\in\mathbb{Z}}\|_2 = \left(\sum_{k\in\mathbb{Z}} |x_k|^2\right)^{\frac{1}{2}}$$

is finite. By Parsevals equality, the norm of $\{x_k\}_{k\in\mathbb{Z}}$ in ℓ_2 equals the L_2-norm of $x(z) = \sum_{k\in\mathbb{Z}} x_k z^k$ on the unit circle $\mathbb{T}$

$$\|\{x_{k\in\mathbb{Z}}\}\|_2^2 = \|x(z)\|_2^2 := \frac{1}{2\pi}\int_{-\pi}^{\pi} |x(e^{i\theta})|^2 d\theta,$$

as it can be shown that $x(z)$ has boundary limits a.e. on $\mathbb{T}$ [34]. Consider the norm which is induced by the input-output behaviour of the system h,

$$\|h\| := \sup_{u(z)\neq 0} \frac{\|y(z)\|_2}{\|u(z)\|_2} = \|h(z)\|_{L_\infty(\mathbb{T})}.$$

The assumption that h is causal imposes the condition that all negative Fourier coefficients of h vanish. If the induced norm of h is finite, then $h(z)$ is analytic in the unit disc $\mathbb{D}$ (see e.g., [32, p. 20]), and the induced norm equals the H_∞-norm

$$\|h\| = \|h(z)\|_{H_\infty}.$$

Such a system h is said to be stable.

One useful concept is that a $f \in H_p$ has an inner-outer factorization, which can be thought of as a "polar form" representation. Given a function $f \in H_p$, it can be factored into two parts, one which is specified by the zeros of f, and one part which is specified by the magnitude of f on the unit circle. The function specified by the zeros of f is called inner, and sometimes referred to as an all-pass function. A function ϕ is inner if it is analytic in $\mathbb{D}$ and $|\phi(z)| = 1$ for a.e. $z \in \mathbb{T}$. Any inner function is of the form $\phi(z) = B(z)S(z)$ where $B(z)$ is a Blaschke product and $S(z)$ is a singular inner function. The Blaschke product and the singular inner function represent two different types of "zeros." A Blaschke product $B(z)$ is of the form

$$B(z) = z^n \prod_{z_\alpha} \frac{|z_\alpha|}{z_\alpha} \frac{z_\alpha - z}{1 - \bar{z}_\alpha z}$$

where n is a positive integer and z_α are zeros in $\mathbb{D}\backslash\{0\}$ such that $\sum_{z_\alpha}(1 - |z_\alpha|) < \infty$, and a singular inner function $S(z)$ is of the form

$$S(z) = \exp\left(-\int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(z)\right)$$

where $d\mu$ is a positive bounded measure on $\mathbb{T}$. The "zero-free" function specified by its magnitude on $\mathbb{T}$ is called an outer function, and is sometimes referred to as a minimum phase function. For the space H_p an outer function F is of the form

$$F(z) = e^{i\gamma} \exp\left(\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} \log(|f(e^{i\theta}|) d\theta \right)$$

where γ is a real number, $\log(|f(e^{i\theta})|) \in L_1$, and $|f(e^{i\theta})| \in L_p$. Now, given a function $f \in H_p$, it has the inner-outer factorization $f(z) = F(z)\phi(z)$, where $F(z)$ is outer and $\phi(z)$ is inner. A more detailed and extensive exposition of Hardy spaces and factorizations can be found in [34, 13].

1.1.2 Degree-constrained analytic interpolation

Given $n+1$ complex numbers $z_0, z_1, \ldots, z_n$ in $\mathbb{D}$ and complex numbers $w_0, w_1, \ldots, w_n$, the classical Nevanlinna-Pick interpolation problem amounts to finding a function f in the *Schur class*

$$\mathcal{S} = \{ f \in H_\infty(\mathbb{D}) \, : \, \|f\|_\infty \leq 1 \}$$

which satisfies the interpolation condition

$$f(z_k) = w_k, \quad k = 0, 1, \ldots, n, \tag{1.1}$$

where H_∞ is the Hardy space of bounded analytic functions on $\mathbb{D}$. Such a function exists if and only if the Pick matrix

$$P = \left[\frac{1 - w_k \bar{w}_\ell}{1 - z_k \bar{z}_\ell} \right]_{k,\ell=0}^{n} \tag{1.2}$$

is positive semidefinite (see, e.g., [16]), and the solution is unique if and only if P is also singular. A complete parameterization of all solutions was given by Nevanlinna (see, e.g., [1]). The Nevanlinna-Pick theory provides one particular solution, which is referred to as the central solution or the maximum entropy solution, since it maximizes the entropy functional

$$\int_{\mathbb{T}} \log(1 - |f|^2) dm$$

subject to (1.1), where dm is the normalized Lebesgue measure on $\mathbb{T}$. This solution is generically of degree n. However, the parameterization given by Nevanlinna provides no knowledge on how to parameterize solutions of low degree.

In several recent papers [6, 7] a theory for parameterizing the low-degree solutions of this problem was developed. The parameterization is in terms of the spectral zeros, and the low-degree solutions are obtained as minimizers of certain entropy functionals. In [8], the problem was solved for positive real interpolants and the theory was further developed in [6, 20]. In [7] the problem was solved for contractive interpolants.

Before presenting this parameterization result, we will introduce some notation. Let ϕ be the finite Blaschke product

$$\phi(z) = \prod_{k=0}^{n} \frac{z_k - z}{1 - \bar{z}_k z}$$

and let $U : f(z) \mapsto zf(z)$ denote the standard shift operator on H_2. Then ϕH_2 is a shift invariant subspace, i.e., $f \in \phi H_2$ implies that $U(f) = zf \in \phi H_2$. Denote by $\mathcal{K}$ the co-invariant subspace $H_2 \ominus \phi H_2$. Then $\mathcal{K}$ is invariant under U^*, where U^* denotes the adjoint of U. Let $\mathcal{K}_0$ denote the set of outer functions in $\mathcal{K}$ that are positive at the origin. The following result is taken from [7].

Theorem 1.1. *Suppose that the Pick matrix* (1.2) *is positive definite, and let σ be an arbitrary function in $\mathcal{K}_0$. Then there exists a unique pair of elements $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ such that*

(i) $f = b/a \in H_\infty$ *with* $\|f\|_\infty \leq 1$,

(ii) $f(z_k) = w_k, \quad k = 0, 1, \ldots, n$, *and*

(iii) $|a|^2 - |b|^2 = |\sigma|^2$ *a.e. on* $\mathbb{T}$.

Conversely, any pair $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ satisfying (i) and (ii) determines, via (iii), a unique $\sigma \in \mathcal{K}_0$. Moreover, setting $\Psi = |\sigma|^2$, the optimization problem

$$\min -\int_{\mathbb{T}} \Psi \log(1 - |f|^2) dm \ \ \text{subject to} \ \ f(z_k) = w_k, \ k = 0, \ldots, n$$

has a unique solution f that is precisely the unique $f \in \mathcal{S}$ satisfying conditions (i), (ii) and (iii).

The zeros of the function $\sigma \in \mathcal{K}_0$, as in the theorem, are the zeros of the spectral outer factor g of the nonnegative function $|g|^2 = 1 - |f|^2$ on the unit circle. Thus, the n zeros of $\sigma(z)^* = \overline{\sigma(\bar{z}^{-1})}$ are often referred to as *spectral zeros*. We also note that the degree of f may be less than n, which happens when a, b, and σ have common roots.

1.1.3 Application to robust control

Robust stabilization is in its most basic setting the problem of finding controllers which stabilize a given set of systems. Examples of real-world problems which can be solved using control are all around us, from control of hard drives in computers and heat systems for houses, to quality assessment for manufacturing plants and technology for airplanes and power plants.

Here we consider a plant P and a controller C with input-output mappings $P : \mathcal{U} \to \mathcal{Y}$ and $C : \mathcal{Y} \to \mathcal{U}$, and assume that P and C are linear time-invariant discrete time systems.

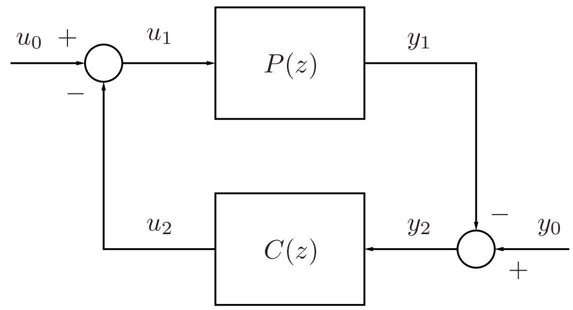

Figure 1.1: Standard feedback configuration.

Figure 1.1 shows a standard feedback configuration. Given signals (u_0, y_0), the inputs (u_1, y_2) and outputs (y_1, u_2) of the plant and controller satisfies

$$\begin{pmatrix} u_0 \\ y_0 \end{pmatrix} = \begin{pmatrix} u_1 \\ y_1 \end{pmatrix} + \begin{pmatrix} u_2 \\ y_2 \end{pmatrix} \tag{1.3}$$

where

$$\begin{aligned} y_1 &= Pu_1 \\ u_2 &= Cy_2. \end{aligned} \tag{1.4}$$

This feedback configuration is internally stable if and only if the mapping (u_0, y_0) to (u_1, u_2, y_1, y_2) is stable and uniquely determined by (1.3) and (1.4). In particular, stability may be expressed in terms of stability of either the mapping from (u_0, y_0) to (u_1, u_2) or the mapping from (u_0, y_0) to (y_1, y_2). These mappings

$$\begin{pmatrix} u_1 \\ y_1 \end{pmatrix} = \Pi_{P//C} \begin{pmatrix} u_0 \\ y_0 \end{pmatrix}, \begin{pmatrix} u_2 \\ y_2 \end{pmatrix} = \Pi_{C//P} \begin{pmatrix} u_0 \\ y_0 \end{pmatrix},$$

can be expressed in terms of the plant and the controller

$$\begin{aligned} \Pi_{P//C} &:= \begin{pmatrix} I \\ P \end{pmatrix} (1 - CP)^{-1}(I, -C) \\ \Pi_{C//P} &:= \begin{pmatrix} C \\ I \end{pmatrix} (1 - PC)^{-1}(-P, I). \end{aligned}$$

The mappings $\Pi_{P//C}$ and $\Pi_{C//P}$ are projections which relate to the gap-robustness of the feedback configuration. For the development and the main results of this framework, see [39, 37, 29, 21, 38].

Stabilization of a feedback system therefore amounts to finding a controller for which the transfer functions in $\Pi_{P//C}, \Pi_{C//P}$ are stable. For the scalar case at hand, one can express this by considering the sensitivity function

$$S(z) = (1 - PC)^{-1}$$

which is the mapping from y_0 to y_2. It turns out that the transfer functions $\Pi_{P//C}, \Pi_{C//P}$ are stable if and only if S is stable and the interpolation conditions

$$\begin{array}{lll} S(z_k) = 1 & \text{whenever} & z_k \text{ is an unstable zero of } P, \text{ and} \\ S(p_k) = 0 & \text{whenever} & p_k \text{ is an unstable pole of } P, \end{array} \tag{1.5}$$

hold [11, 12]. These interpolation conditions reflect that there can not be any unstable cancellation of poles or zeros in the product PC.

Problems of robust control address the effect of uncertainties in the nominal plant model and the effect of external disturbances. Typical models for plant uncertainty are taken in the form of additive or multiplicative perturbations of the nominal transfer function, perturbations of a coprime factor representation, or in a suitable metric topology such as the one induced by the gap metric. In such cases where the description of uncertainty is "unstructured," i.e., non-parametric, the robustness margin and the performance of the closed loop system can often be expressed in terms of the norm or the weighted norm of $\Pi_{P//C}$ or its various entries [11, 37, 21]. Indeed, stabilization may be cast as an interpolation problem with an analytic function [11], e.g., when the sensitivity function is required to be stable and to satisfy (1.5) in order to ensure internal stability of the closed loop system. The performance of the closed loop system relates to the shape of the sensitivity function which relates to disturbance rejection and tracking at various frequency ranges. Robustness to non-parametric perturbations relates again, via the small gain theorem, to the shape of the sensitivity function or other suitable closed-loop mappings depending on the uncertainty description. In either case, such problems amount to seeking a small (weighted) norm for an analytic function that abides by given interpolation constraints. The solution in the form of an analytic interpolant, e.g., a suitable sensitivity function, allows recovering the corresponding stabilizing controller C.

Naturally, the degree of such an interpolant relates to the degree of the corresponding controller. For instance, in the context of sensitivity shaping where a sensitivity function $S = 1/(1 - PC)$ is sought that satisfies suitable constraints, the degree of the controller $C = (S - 1)/PS$ is bounded by

$$\deg C \leq \deg P + \deg S - n_z + n_p$$

where n_z and n_p are the number of unstable zeros and poles of the plant, respectively [30]. In a similar manner, robustness in coprime factor uncertainty or, uncertainty in the gap metric, may also be posed as an interpolation problem and an analogous degree bound for the controller holds. Interest in keeping the degree of C, S or other input-to-error maps small can be motivated from several different angles. First, the complexity of the controller adds to numerical issues, round-off errors, implementation delays, and computational delays. Second, a closed loop system may be the subject of control via an outer loop that may include a human operator. Naturally, a high order input-to-error map will create an extra impediment in controlling the system by providing suitable inputs for needed tasks. In conclusion, it is highly desirable to achieve performance and robustness while maintaining the degree of interpolants/C within reasonable bounds. This issue will be studied in detail in Paper A. In the first part of this thesis we derive methods for shaping and approximation of interpolants, and the design of low degree controllers.

1.2 Stochastic processes and spectral estimation

The power spectrum of a time-series signal represents the energy content over frequencies of the signal. Spectral estimation can therefore be thought of as estimation of the frequency content of a signal. This is a widely used methodology, which is used for analysis and prediction in areas such as speech, stock market, radar, sonar, and several medical applications (see e.g., [36, 25, 3]).

Consider a discrete-time stochastic process, which is represented by a mapping from a probability space to a time series

$$\ldots, y_{-1}, y_0, y_1, \ldots.$$

In addition we assume that this stochastic process has zero mean, i.e.,

$$\mathcal{E}(y_k) = 0 \text{ for all } k \in \mathbb{Z},$$

and is second order stationary, i.e.,

$$\mathcal{E}(y_t \bar{y}_{t-k}) \text{ is independent of } t,$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator. Stationarity allows us to define the covariances as $c_k = \mathcal{E}(y_t \bar{y}_{t-k})$. The power spectrum $d\mu$ of the process is the positive measure on $\mathbb{T} = \{z : |z| = 1\}$ with covariances as Fourier coefficients

$$c_k := \int_{-\pi}^{\pi} e^{-ik\theta} d\mu(e^{i\theta}) \text{ for } k \in \mathbb{Z}.$$

If the spectrum is smooth, then the power spectral density of the time series is

$$\Phi(e^{i\theta}) = \sum_{k=-\infty}^{\infty} c_k e^{ik\theta},$$

where $\Phi(e^{i\theta})d\theta = d\mu(e^{i\theta})$.

For any power spectrum $d\mu$ with covariances $(c_0, c_1, c_2, \ldots)$, there is an associated analytic function

$$f(z) = \frac{1}{2}c_0 + c_1 z + c_2 z^2 \ldots.$$

This function satisfies $2\Re e f(e^{i\theta})d\theta = d\mu(e^{i\theta})$ a.e., or in the case with a smooth spectrum $2\Re e f(e^{i\theta}) = \Phi(e^{i\theta})$ holds. According to Herglotz' theorem [2] this function satisfies

$$f(z) = \frac{1}{4\pi} \int_{-\pi}^{\pi} \frac{e^{i\theta} + z}{e^{i\theta} - z} d\mu(e^{i\theta}) \tag{1.6}$$

and is analytic in $\mathbb{D} := \{z : |z| < 1\}$ with nonnegative real part. Any analytic function in $\mathbb{D}$ with nonnegative real part is referred to as either "positive-real" or, as a Carathèodory function.

Consider the inverse problem to determine when there exists a power spectrum with covariance sequence $(c_0, c_1, \ldots, c_n)$, and if such power spectrum exists, to parameterize all of them. This is in fact an interpolation problem where a positive real $f(z)$ is sought such that

$$f(0) = \frac{1}{2}c_0, \quad \frac{f^k(0)}{k!} = c_k \text{ for } k = 1, \ldots, n. \tag{1.7}$$

If the Toeplitz matrix corresponding to this interpolation problem,

$$T = \begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-n} \\ c_1 & c_0 & \cdots & c_{-n+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_n & c_{n-1} & \cdots & c_0 \end{bmatrix}, \tag{1.8}$$

is positive definite, then there is a family of infinitely many power spectra consistent with the covariance sequence [23]. If the Toeplitz matrix (1.8) is positive semidefinite and singular, then there is a uniquely defined power spectrum $d\mu$ consistent with the covariance sequence.

When a stochastic process $\{u_k\}_{k\in\mathbb{Z}}$ with power spectrum $d\mu$ is passed through a LTI filter h, the power spectrum of the output is $\left|h(e^{i\theta})\right|^2 d\mu(e^{i\theta})$. Therefore, the stochastic process $\{y_k\}_{k\in\mathbb{Z}}$, produced by passing white noise through an ARMA filter σ/a, has the power spectrum

$$\Phi(e^{i\theta})d\theta = \left|\frac{\sigma(e^{i\theta})}{a(e^{i\theta})}\right|^2 d\theta.$$

Conversely, if Φ is rational, then it may be factored into $h(z)h(z)^*$, where $h(z) = \sigma(z)/a(z)$ is the transfer function of an ARMA-system. The corresponding positive real function f then satisfies

$$f(z) + f(z)^* = \Phi(z) = h(z)h(z)^*, \text{ for } z \in \mathbb{T},$$

and hence $\sigma(z)$ is determined by $\sigma(z)\sigma(z)^* = a(z)b(z)^* + a(z)^*b(z)$. The zeros of $\sigma(z)^*$ are referred to as the *spectral zeros* of the positive real function $f(z)$.

1.2.1 Identification and spectral estimates

In most practical applications, we study a stochastic process $\{y_k\}_{k\in\mathbb{Z}}$, assumed to be zero-mean, second-order stationary, ergodic, and with power spectrum $d\mu$. Spectral estimation is to find an estimate $d\hat{\mu}$ of $d\mu$ based on the finite sample $y_1, \ldots, y_N$.

Here, the estimation of a power spectrum is done in two steps. First estimating covariances c_k from the sample $y_1, \ldots, y_N$, and then finding a power spectrum which has the estimated covariances as Fourier coefficients. The covariances can be estimated using the averages

$$\hat{c}_k = \frac{1}{N}\sum_{\ell=k+1}^{N} y_\ell \bar{y}_{\ell-k}, \tag{1.9}$$

which are referred to as the biased covariance estimates. Ergodicity of the process ensures that $\hat{c}_k$ is a good estimate of c_k as long as $k \ll N$, and that $\hat{c}_k$ converges to c_k almost surely as N tends to infinity. For any such estimate $(\hat{c}_0, \hat{c}_1, \ldots, \hat{c}_n)$, the Toeplitz matrix (1.8) is positive. This ensures that there exist a power spectrum $d\hat{\mu}$ that matches the covariances

$$\hat{c}_k = \int_{-\pi}^{\pi} e^{-ik\theta} d\hat{\mu}(e^{i\theta}) \text{ for } k = 0, \ldots, n.$$

In general there is an infinite family of such power spectrum, and below we will describe two methodologies for choosing specific power spectra.

The periodogram is the spectral estimate

$$\Phi_P(e^{i\theta}) = \frac{1}{N} \left| \sum_{\ell=1}^{N} y_\ell e^{-i\theta\ell} \right|^2$$

of the spectral density Φ [35]. It identifies periodicities in the signal by multiplying the signal by a complex exponential. It can be shown that this equals

$$\Phi_P(e^{i\theta}) = \sum_{k=-N+1}^{N-1} \hat{c}_k e^{ik\theta}$$

where $\hat{c}_k$ are the biased covariance estimates from (1.9) [4]. The main objection to using the periodogram in spectral estimation is that it has poor pointwise convergence properties [36]. To deal with this, smoothing and/or windowing may be utilized. In this text, we consider the problem from a different viewpoint, and instead of changing the spectral estimate, the topology is changed to one where the periodogram converges nicely.

The second methodology is based on finding a rational spectrum of low degree that match the covariances. Given a positive covariance sequence $(c_0, c_1, \ldots, c_n)$, there is always an AR-filter a of degree n such that the power spectrum $\frac{1}{|a(e^{i\theta})|^2} d\theta$ has Fourier coefficients equal to the given covariance. It turns out that this is the power spectrum with the highest entropy gain

$$\int_{\mathbb{T}} \log\left(\frac{d\mu}{d\theta}\right) dm$$

that satisfies the covariance constraint [5]. Furthermore, this solution can be found by solving a set of linear equations, and it is of the form $d\mu = \Phi d\theta$ where Φ is of degree $2n$.

In [26], R.E. Kalman put forth a rational covariance extension problem, which in essence amounts to parameterizing all rational functions Φ of degree bounded by $2n$ such that $\Phi d\theta$ is a positive measure with covariances $(c_0, c_1, \ldots, c_n)$. It was proved in [18, 17] that for every given MA-filter σ, there is an AR-filter $\frac{1}{a}$ such that

$$d\mu(e^{i\theta}) = \left| \frac{\sigma(e^{i\theta})}{a(e^{i\theta})} \right|^2 d\theta$$

is consistent with $(c_0, c_1, \ldots, c_n)$, and it was conjectured that there is a unique such AR-filter with $a(0) > 0$. The conjecture was established in [9, 10, 19] and a constructive method was developed in [8] for calculating solutions via a convex optimization problem. This is described in the following theorem.

Theorem 1.2. *Let $(c_0, c_1, \dots, c_n)$ be a covariance sequence for which the Toeplitz matrix* (1.8) *is positive definite, and let σ be an arbitrary non-zero polynomial of degree n. Then there exists a unique stable polynomial a of degree n with $a(0) > 0$ such that*

(i) $d\mu(e^{i\theta}) = \left|\frac{\sigma(e^{i\theta})}{a(e^{i\theta})}\right|^2 d\theta$ *is a positive measure and*

(ii) $c_k = \int_{-\pi}^{\pi} e^{-ik\theta} d\mu(e^{i\theta}) \quad k = 0, 1, \dots, n.$

Moreover, setting $\Psi = |\sigma|^2$, the optimization problem

$$\max \int_{-\pi}^{\pi} \Psi(e^{i\theta}) \log\left(\frac{d\mu(e^{i\theta})}{d\theta}\right) dm \;\; \textit{subject to} \;\; c_k = \int_{-\pi}^{\pi} e^{-ik\theta} d\mu(e^{i\theta}),\ k = 0, \dots, n$$

has a unique solution $d\mu$ that is precisely the unique $d\mu$ satisfying conditions (i), (ii).

This theory has been further developed for general weighting functions Ψ in [27]. Approximation of positive real interpolants using approximation of the corresponding weights has been developed in [15, 14]. This follows a similar rationale as used in Paper A and Paper C.

The function f corresponding to $d\mu$ according to (1.6) satisfies the interpolation conditions (1.7). This may be used for characterizing all positive real interpolants of degree bounded by n. In fact, the set of positive real functions $f = b/a$ of degree bounded by n satisfying (1.7), may be parameterized in terms of σ, where $ab^* + a^*b = \sigma\sigma^*$ [8].

1.2.2 Weak*-convergence

The more data from the sequence $\{y_k\}_{k\in\mathbb{Z}}$ is used, the better estimates of the underlying power spectrum can be obtained. Here we consider the spectral estimation problem based on the sample

$$y_1, \dots, y_N$$

and convergence properties of the spectral estimates as N tends to infinity.

Let $\Phi_{P,N}(e^{i\theta})$ be the periodogram based on the sequence $y_1, \dots, y_N$. If the true underlying power spectra is $d\mu(e^{i\theta}) = \Phi(e^{i\theta})d\theta$, then the periodogram is an unbiased estimator of Φ, i.e.,

$$\lim_{N\to\infty} \mathcal{E}(\Phi_{P,N}(e^{i\theta})) = \Phi(e^{i\theta}).$$

However, the asymptotic variance is

$$\lim_{N\to\infty} \mathcal{E}\left(\Phi_{P,N}(e^{i\theta}) - \Phi(e^{i\theta})\right)^2 = \Phi(e^{i\theta})^2,$$

and hence the variance of the error does not go to zero as $N \to \infty$ (see e.g., [36]). Therefore, the periodogram is not asymptotically efficient as a pointwise estimator of the power spectral density.

The maximum entropy spectrum has nicer pointwise convergence properties, but convergence can only be guaranteed if the power spectrum is sufficiently smooth. Y. L. Geronimus [22] proved that the maximum entropy spectrum converges uniformly under such assumptions. In practice, a priori information about such smoothness is rarely available. Therefore, we are led to consider a weaker type of convergence where only localization of the spectral power is required. This is explained next.

We denote by $C(\mathbb{T})$ the class of real-valued continuous functions on $\mathbb{T}$. The set of measures on $\mathbb{T}$ can be identified with the dual space of $C(\mathbb{T})$, i.e., the space of bounded linear functionals $\Lambda : C(\mathbb{T}) \to \mathbb{R}$ ([24], [28]). This is the Riesz representation theorem, which asserts the existence of a bounded measure $d\mu$ such that

$$\Lambda(G) = \int_{\mathbb{T}} G(z) d\mu(z)$$

for all $G \in C(\mathbb{T})$. Thus, for any two measures that are different, there exists a continuous function such that the two measures integrate to different values. In other words, continuous functions serve as "test functions" to differentiate between measures, and bounds on such integrals define the weak*-topology. A sequence of measures $d\mu_n$, $n = 1, 2, \ldots$, converges to $d\mu$ in the weak* topology if

$$\int_{\mathbb{T}} G d\mu_n \to \int_{\mathbb{T}} G d\mu \text{ for every } G \in C(\mathbb{T}).$$

The limit, when it exists, is defined uniquely by the sequence.

The weak*-topology on the set of power spectra respects the statistical properties of the signal and localization of power spectra. The periodogram converges to the true spectrum in the weak* topology. This was observed in [33] (cf. [24, page 24]).

Theorem 1.3. *Let $\Phi_{P,N}$ be the periodogram obtained from the sample $(y_0, y_1, \ldots, y_N)$ of a stochastic process with power spectrum $d\mu$. Then*

$$\Phi_{P,N} d\theta \to d\mu \text{ in weak}^*$$

almost surely.

In fact, any sequence of spectral estimators $d\mu_n$ with covariances $(c_0^n, c_1^n, \ldots)$ such that $c_k^n \to c_k$ for all k as $n \to \infty$ converges to $d\mu$ in weak*. Therefore, the maximum entropy estimate also converges to the true spectrum as $n \to \infty$ in weak*.

1.2.3 Application to speech processing

Spoken communication between people consists of sound waves created in the vocal tract, which transmits messages. The sound wave excites the tympanic membrane in the ear which passes the information on to the brain. In speech analysis each part of this communication is analyzed in order to understand a message digitally and in speech synthesis sound waves are produced which resembles specific messages and phrases.

Sampling a speech waveform gives rise to a time series representing the acoustic sound pressure wave. The data carries information about both the speaker and the message. There are many methods to analyze the data and obtain useful information from the speech sample, and most of the methods use local second order statistics. There are two main reasons for this. Firstly, due to the physiology of the human speech production system, the speech signal may be modelled as the output of a second order stationary process for short time periods. Secondly, the second order statistics characterize the power of the signal over the frequencies, a characteristic of the sound wave which the human auditory system is sensitive to.

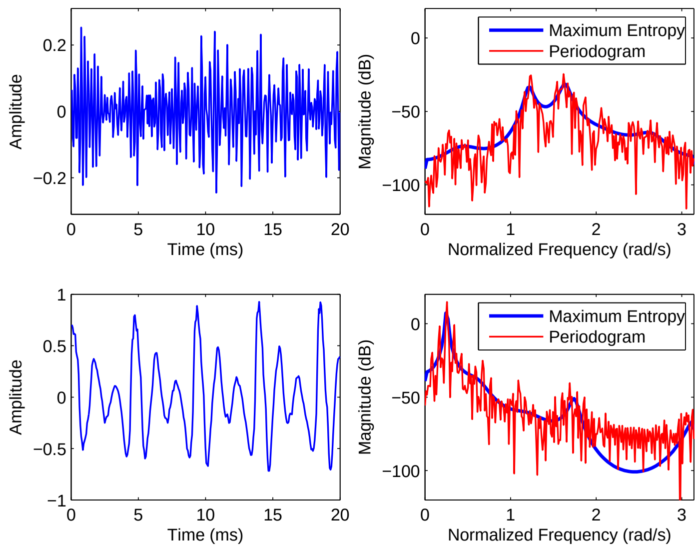

Figure 1.2: Time-domain speech signal and power spectral estimates of the phonemes $/S/$ and $/A/$ respectively.

One example where this is used in everyday life is in a cellphone. Since there are limits in the amount of data which may be sent through the telecommunication network, the entire time series representing the sound wave can not be sent. Instead, the sampled speech is divided into short segments and for each segment, the maximum entropy spectrum is calculated from the covariances. The spectrum is then sent together with an input sequence which carries information about the excitation and periodicity of the speech wave. Consequently, the speech produced by the cellphone is artificially recreated, but nevertheless it resembles the important characteristics of the sounds.

Phonemes are small segments of speech, and a sound wave can be thought of as a sequence of phonemes. In speech analysis, a common procedure is to identify phonemes based on their power spectra. Consider two speech waves sampled from the phonemes $/S/$ and $/A/$, which are depicted on the left in Figure 1.2. For identifying which phonemes the speech waves corresponds to, their spectra are estimated, and the power spectral estimates corresponding to the maximum entropy spectrum and the periodogram are shown on the right. The spectra obtained are then compared to a database, where different spectra corresponding to the phonemes are stored (usually utilizing hidden markov models). The main theme of the second part of this thesis deals with how to construct distances that are appropriate for such discrimination and which properties such distances should have.

1.3 Main contributions of this thesis

This thesis is divided into two parts. The first part, containing Papers A-C, deals with approximation subject to interpolation constraints. The second part, containing Papers D-E, deals with spectral estimation and weak*-continuous distances. The main results are explained below.

1.3.1 Approximation subject to interpolation constraints

In [7] the degree-constrained Nevanlinna-Pick problem was solved for contractive interpolants, and this circle of ideas is studied in detail in Paper A and Paper B. The parameterization of the interpolants is done in terms of minimizers of

$$\min -\int_{\mathbb{T}} \Psi \log(1-|f|^2)dm \ \text{ subject to } \ f(z_k)=w_k,\ k=0,\dots,n \tag{1.10}$$

where the weighted entropy functional is to be minimized. Whenever $\Psi=|\sigma|^2$ for some $\sigma\in\mathcal{K}_0$, the degree of the minimizer is bounded by n. In Paper A, we generalize this result for handling arbitrary log-integrable weights Ψ. As a corollary to this theorem, the set of weights $\mathcal{K}_0$ may be extended in a natural way for parameterizing interpolants of degree bounded by $n+m$ for any given $m\geq 0$. As a second corollary to the theorem, we obtain a solution to the following inverse problem: given an interpolant f, determine the class of weights Ψ so that the minimizer of (1.10) is the given function f.

Using these results we develop a degree reduction procedure for interpolants. This gives a solution to the previously partly open problem on how to choose the weight Ψ (i.e., spectral zeros of f) in order to obtain a particular shape. Given a interpolant g of high degree, first find a weight $|\rho|^2$ so that g is the minimizer of (1.10) using weight $\Psi=|\rho|^2$. Then use quasi-convex optimization to find the $\sigma\in\mathcal{K}$ such that $|\sigma|^2$ is as close as possible to $|\rho|^2$. Then, letting f be the minimizer of (1.10) using the weight $|\sigma|^2$, gives a interpolant f that is of low degree and close to g. By approximating in the weights instead of directly in the interpolants, a nonconvex problem is replaced by one that is tractable. Bounds on the approximation error, $f-g$, are given in terms of the optimum of the quasi-convex optimization problem.

We consider two control applications, one dealing with sensitivity shaping and the other dealing with frequency-dependent robustness in the coprime factors of the plant.

Both problems can be formulated as interpolation problems and degree bounds on the controller can be obtained based on the degree of the interpolant. The advantage of using our method, compared to standard H_∞ theory, is that it allows for shaping interpolants without increasing the degree unduly.

In Paper B the role of weights Ψ and interpolation points for shaping the minimizer f of the minimization problem (1.10) is considered. It is shown that, if, for a sequence of interpolants, the values of the entropy gain of the interpolants converge to the optimum, then the interpolants converge in H_2, but not necessarily in H_∞. This result is then used to describe the asymptotic behaviour of the interpolant as an interpolation point approaches the unit circle. The result is, under a condition on the weight, that adding an interpolation point close to the unit circle have negligible effect on the H_2-norm, but it may have large effect on the H_∞ norm.

For loop shaping to specifications in control design, it might at first seem natural to place strategically additional interpolation points close to the boundary. However, our results indicate that such a strategy will have little effect on the shape. Another consequence of our results relates to model reduction based on minimum-entropy principles, where one should avoid placing interpolation points too close to the boundary. We have analyzed a design example from robust control, studied by Nagamune [31], in the context of our results. It turns out that the effect of an added interpolation condition close to the boundary turns out to be small, but that the effect of changing the weight is major.

These results are also utilized in Paper A for studying the continuity of the mapping from weights Ψ to minimizers f of the minimization problem (1.10). It is shown that the mapping from $\log \Psi \in L_\infty$ to H_2 is continuous, but the mapping from $\log \Psi \in L_\infty$ to H_∞ is not. The former statement is in fact used above for deriving the error bound in the approximation procedure.

By scaling the interpolant $f \to \gamma^{-1} f$, the optimization problem (1.10) becomes

$$\min -\gamma^2 \int_{\mathbb{T}} \Psi \log(1-\gamma^{-2}|f|^2) dm \quad \text{subject to} \quad f(z_k) = w_k, \ k = 0, \ldots, n$$

and can be used to parameterize interpolants with $\|f\|_\infty \leq \gamma$. In Paper C we study what happens as $\gamma \to \infty$. It turns out that

$$-\gamma^2 \int_{\mathbb{T}} \Psi \log(1-\gamma^{-2}|f|^2) dm \to \int_{\mathbb{T}} \Psi |f|^2 dm \text{ as } \gamma \to \infty,$$

i.e., that the entropy functional converges to a weighted H_2-norm as $\gamma \to \infty$. This leads to a parameterization result similar to the ones obtained for contractive interpolation and positive real interpolation. The functions $\sigma \in \mathcal{K}_0$ parameterize the analytic functions $f = \frac{b}{a}$ where $b \in \mathcal{K}, a \in \mathcal{K}_0$ in terms of minimizers of the entropy functionals in the following table.

Property		Entropy functional	Spectral zeros
Passive	$\Re e(f) \geq 0$	$\int \lvert\sigma\rvert^2 \log \Re e(f)$	$\sigma\sigma^* = a^*b + ab^*$
Bounded	$\lVert f\rVert_\infty \leq \gamma$	$-\gamma^2 \int \lvert\sigma\rvert^2 \log(1-\gamma^{-2}\lvert f\rvert^2)$	$\sigma\sigma^* = aa^* - \gamma^{-2}bb^*$
Stable	$f \in H(\mathbb{D})$	$\int \lvert\sigma f\rvert^2$	$\sigma\sigma^* = aa^*$

In Paper C, we also develop a procedure similar to the one derived in Paper A for approximation of interpolants. Also here, the basic idea is to use the inverse problem and quasi-convex optimization to find a weight for which the minimizing interpolant is of both low degree and attains a desired shape. In some problems, only a few interpolation conditions are given. This gives an additional freedom of choosing interpolation points as extra design parameters. By considering the error bound of the approximation, the interpolation points may be chosen in a systematic way in order to guarantee good approximation in selected regions.

1.3.2 Spectral estimation and weak*-continuous distances

In the second part, we consider metrics and topologies for power spectra. In Paper D, we start by studying spectral estimation based on covariances. Consider a positive measure $d\nu$ with covariance sequence $\mathbf{c} = (c_0, c_1, \ldots)$ and consider the set of power spectra consistent with the n first covariances

$$\mathcal{F}_{\mathbf{c}_{0:n}} = \left\{ d\mu \geq 0 : c_k = \int_{-\pi}^{\pi} e^{-ik\theta} d\mu, k = 0, 1, \ldots, n \right\}.$$

Identification often builds on covariance estimates, and if only the finite covariance sequence $(c_0, c_1, \ldots, c_n)$ is given, the only thing known about the measure $d\nu$ is that it belong to $\mathcal{F}_{\mathbf{c}_{0:n}}$. As n tends to ∞, this set shrinks and in the limit the only measure consistent with the entire covariance sequence $\mathbf{c}$ is $d\nu$

$$\bigcap_{n \in \mathbb{N}} \mathcal{F}_{\mathbf{c}_{0:n}} = \{d\nu\}.$$

Therefore one would expect that the diameter of $\mathcal{F}_{\mathbf{c}_{0:n}}$ with respect to the distances δ,

$$\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}}) := \sup\{\delta(d\mu_0, d\mu_1) \,:\, d\mu_0, d\mu_1 \in \mathcal{F}_{\mathbf{c}_{0:n}}\},$$

also shrinks nicely as $n \to \infty$. This is not the case for most of the standard distance functions used for discriminating between power spectra. If δ is taken as the Itakura-Saito distance, the Kullback-Leibler distance, or the logaritmic distance, then the diameter is infinite for any n. If the total variation is considered, the diameter $\rho_{\delta_{\text{TV}}}(\mathcal{F}_{\mathbf{c}_{0:n}})$ is $2c_0$ for any n. The problem is that, determining the values of the spectral density at a point based on a finite set of covariances is an ill-posed problem, and all the distance functions above are based on the pointwise behaviour of the spectra.

However, we prove that, if δ is weak*-continuous then the diameter $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ shrinks to 0 as $n \to \infty$. By using weak*-continuous distances, localization of the measure is possible, i.e., determining the value of

$$\int_{\mathbb{T}} G(z) d\nu(z)$$

with increasing accuracy as n increases, whenever G is a continuous function. Motivated by this, several ways of constructing weak*-continuous metrics are proposed. For one class of such metrics, explicit bounds on the diameter of $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ are given in terms of the covariance sequence. Finally, we consider an example where the trade-off between variance and resolution is considered. In this setting, the choice of metric represents the required resolution, and then sufficient data (covariances) need to be provided guarantee that the variance is low.

In Paper E, we consider distances for power spectra based on the Monge-Kantorovic transportation problem. One characteristic of a good distance measure is that is has a set of useful properties. In this paper we consider the inverse problem of finding distances which satisfy a set of natural properties. This axiomatic approach builds on four axioms. We require that the distance is 1) a metric, 2) contractive with respect to additive noise, 3) contractive with respect to multiplicative noise, and 4) weak*-continuous. When signals are corrupted with noise, discrimination between signals becomes more difficult. This is the motivation for the contractiveness properties. The weak*-continuity is required in order to ensure that the metric is continuous with respect to the second order statistics. We contrast this by considering earlier established distances based on information theory and statistics. In this context, another property which seem natural is invariance with respect to linear filtering. However, we show that this property is incompatible with any weak*-continuous metric.

The Monge-Kantorovic transportation problem is a problem, where the cost of transporting mass from one measure $d\mu_0$ to another measure $d\mu_1$ is to be minimized. The Wasserstein distance is a metric, associated with the minimal transportation cost, which satisfies several of the requested properties. However, it only allows for differentiating between measures of equal mass. To handle the situations with possibly unequal mass, we postulate that the measures $d\mu_0, d\mu_1$ are perturbations of two other measures $d\nu_0, d\nu_1$ which have equal mass. Then, the cost of transporting the original measures $d\mu_0, d\mu_1$ to one another can be thought of as the cost of transporting the perturbed measures $d\nu_0, d\nu_1$ to one another plus the size of the respective perturbations. This leads to the optimization problem

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) := \inf_{\int d\nu_0 = \int d\nu_1} T_c(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^{1} \delta_{\text{TV}}(d\mu_i, d\nu_i),$$

where T_c is the transportation cost and δ_{TV} is the total variation. By using the cost function

$c(x,y) = |x-y|^p$, the distance function

$$\delta_{p,\kappa}(d\mu_0, d\mu_1) := \left(\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)\right)^{\min(1,\frac{1}{p})}$$

imposes a metric on the set of power spectra which satisfies all the above properties. At first, computation of the distance appears to be a problem. However, by considering the dual formulation, this may be posed as the linear optimization problem

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \mathcal{J}(\phi,\psi)$$

where

$$\mathcal{J}(\phi,\psi) = \int \phi d\mu_0 + \psi d\mu_1.$$

and

$$\begin{aligned}\Phi_{c,\kappa} \quad := \quad & \big\{(\phi,\psi) \in L^1(d\mu_0) \times L^1(d\mu_1) \,:\, \phi(x) \le \kappa, \\ & \psi(y) \le \kappa\,,\, \phi(x) + \psi(y) \le c(x,y)\big\}.\end{aligned}$$

Therefore, it may be calculated using standard methods for solving linear optimization problems. One additional property of the transport distances is that their geodesics represent smooth transitions of mass. This is an important contrast to standard distances used for power spectra which are mainly based on the pointwise distance between the spectra on the unit circle.

1.4 Summary of the papers

This thesis consists of five papers, and hereafter follows a short summary of each of them.

A: **The Inverse Problem of Analytic Interpolation with Degree Constraint and Weight Selection for Control Synthesis.** The paper is coauthored with Tryphon Georgiou and Anders Lindquist and has been provisionally accepted for publication in *IEEE Transactions of Automatic Control*. Results from this paper have appeared or will appear in the proceedings of CDC 2006, MTNS 2008, CDC 2008.
The minimizers of certain weighted entropy functionals are the solutions to an analytic interpolation problem with a degree constraint, and all solutions to this interpolation problem arise in this way by a suitable choice of weights. This allows for parameterizing the set of contractive interpolants with a degree bound by selecting weights. An approximation procedure is developed as follows. First the set of weights which give rise to an interpolant of high degree is found. This is referred to as the inverse problem. Then based on an optimization problem, the weight which give rise to a low degree interpolant and is closest to the former weight is found. This weight is used to obtain an interpolant of low degree, and which is close to the interpolant of high degree. Instead of approximation of the interpolant directly, which is

a non-convex problem, the problem of approximating the weight is tractable.
This approximation procedure is used in order to systematize feedback control synthesis with a constraint on the degree. This is utilized for weighted sensitivity shaping and for design of controllers for weighted robustness in the coprime factors, while at the same time keeping the degree of the controller and the closed loop low.

B: **On Degree-Constrained Analytic Interpolation with Interpolation Points Close to the Boundary.** The paper is coauthored with Anders Lindquist and has been provisionally accepted for publication in *IEEE Transactions of Automatic Control*. Results from this paper has appeared in the proceedings of MTNS 2006.
In this paper we study contractive interpolants obtained as minimizers of a weighted entropy functional, and analyze the role of weights and interpolation conditions as design parameters for shaping the interpolants. We first show that, if, for a sequence of interpolants, the values of the entropy gain of the interpolants converge to the optimum, then the interpolants converge in H_2, but not in H_∞. This result is then used to describe the asymptotic behaviour of the interpolant as an interpolation point approaches the boundary of the domain of analyticity. For loop shaping to specifications in control design, it might at first seem natural to place strategically additional interpolation points close to the boundary. However, our results indicate that such a strategy will have little effect on the shape. Another consequence of our results relates to model reduction based on minimum-entropy principles, where one should avoid placing interpolation points too close to the boundary. We have analyzed a design example from robust control, studied by Nagamune [31], in the context of our results. It turns out that the effect of an added interpolation condition close to the boundary is small, but that the effect of changing the weight is major.

C: **Stability-Preserving Rational Approximation Subject to Interpolation Constraints.** The paper is coautored with Anders Lindquist and has been accepted for publication in *IEEE Transactions of Automatic Control*. The conference version of this paper for CDC 2007 was a finalist for the best student paper award.
This paper presents new theory for stability-preserving model reduction that can also handle prespecified interpolation conditions and comes with error bounds. We have presented a systematic optimization procedure for choosing an appropriate weight so that the minimizer of a corresponding weighted H_2 minimization problem both matches the original system and has low degree.
The study of the H_2 minimization problem is motivated by the relation between the H_2 norm and the entropy functional used in the bounded interpolation problem. The H_2 minimization problem may be seen as the limit case of bounded interpolation, but where the $\infty-$norm bound is removed and only stability is required. Therefore, new concepts derived in this framework are useful for understanding entropy minimization.

D: **Localization of Power Spectra.** The paper is coauthored with Tryphon Georgiou. Results from this paper has appeared in the proceedings of CDC 2005.
In this paper we study the problem of robust identification of power spectra based

on covariance measurements and the purpose is to study the topology and develop metrics that allows for localization of power spectra. The main conclusion is that, for our robustness criteria to be satisfied with respect to a metric, the metric needs to be weak*-continuous. Several examples on how to construct such metrics are given. This allows us to quantify uncertainty of spectra in a natural way and to calculate a priori bounds on spectral uncertainty, based on second-order statistics. Finally, we study identification of spectral densities and relate this to the trade-off between resolution and variance of estimates.

E: **Metrics for Power Spectra: an Axiomatic Approach.** This paper is coauthored with Tryphon Georgiou and Mir Shahrouz Takyar and has been accepted for publication in *IEEE Transactions of Signal Processing*. Results from this paper will appear in the proceedings of CDC 2008.
In this paper we propose an axiomatic way of determining metrics for power spectra. A set of properties are put forth, which requires the metric to be 1) continuous with respect to second order statistics, and 2) contractive with respect to certain transformations which correspond to corruption of the signal. We then present a particular metric which abides by these requirements. The metric is based on the Monge-Kantorovich transportation problem and is contrasted to an earlier Riemannian metric based on the minimum-variance prediction geometry of the underlying time-series. Finally we compare the distance with the Itakura-Saito distance and a metric based on prediction, in two identification examples.

1.5 Bibliography

[1] J. Agler and J.E. McCarthy. *Pick Interpolation and Hilbert Function Spaces*. American Mathematical Society, 2002.

[2] N. I. Akhiezer. *The classical moment problem*. Hafner Publishing, Translation 1965. Oliver and Boyd.

[3] A.N. Amini, E. Ebbini, and T.T. Georgiou. Noninvasive estimation of tissue temperature via high-resolution spectral analysis techniques. *IEEE Trans. on Biomedical Engineering*, 52(2):221–228, February 2005.

[4] R.B. Blackman and J.W. Tukey. *The Measurement of Power Spectra, from the Point of View of Communications Engineering*. Dover, 1959.

[5] J. Burg. *Maximum Entropy Spectral Analysis*. PhD thesis, Stanford University, 1975.

[6] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[7] C. I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski. Generalized interpolation in H-infinity with a complexity constraint. *Transactions of the American Mathematical Society*, 358:965–987, 2006.

[8] C. I. Byrnes, S. V. Gusev, and A. Lindquist. A convex optimization approach to the rational covariance extension problem. *SIAM J. Contr. and Optimiz.*, 37:211–229, 1998.

[9] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev. A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Automat. Control*, 40:1841–1857, 1995.

[10] C. I. Byrnes, A. Lindquist, and H. J. Landau. On the well-posedness of the rational covariance extension problem. In M. Alber, B. Hu, and J Rosenthal, editors, *Current and Future Directions in Applied Mathematics*, pages 81 –108. Birkhäuser Boston, 1997.

[11] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory*. Macmillan Publishing Company, New York, 1992.

[12] G. E. Dullerud and F. G. Paganini. *A Course in Robust Control Theory: A Convex Approach*. Springer, 2000.

[13] P.L. Duren. *Theory of H^P spaces*. Dover, 2000.

[14] G. Fanizza. *Modeling and Model Reduction by Analytic Interpolation and Optimization*. PhD thesis, Optimization and System Theory, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 2008.

[15] G. Fanizza, J. Karlsson, A. Lindquist, and R. Nagamune. Passivity-preserving model reduction by analytic interpolation. *Linear Algebra and its Applications*, 425:608–633, 2007.

[16] J.B. Garnett. *Bounded analytic Functions*. Academic Press, 1981.

[17] T. T. Georgiou. A topological approach to Nevanlinna-Pick interpolation. *SIAM J. on Mathematical Analysis*, 18(5):1248–1260, September 1987.

[18] T.T. Georgiou. *Partial Realization of Covariance Sequences*. PhD thesis, CMST, University of Florida, Gainesville, 1983.

[19] T.T. Georgiou. The interpolation problem with a degree constraint. *IEEE Trans. on Automatic Control*, 44:631–635, March 1999.

[20] T.T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. on Information Theory*, 49(11), November 2003.

[21] T.T. Georgiou and M. Smith. Optimal robustness in the gap metric. *IEEE Trans. on Automatic Control*, 35(6), June 1990.

[22] Ya. L. Geronimus. *Orthogonal Polynomials*. Consultants Bureau, 1961.

[23] U. Grenander and G. Szegö. *Toeplitz Forms and their Applications*. Chelsea, 1958.

[24] K. Hoffman. *Banach Spaces of Analytic Functions*. Dover Publications, 1962.

[25] J.R. Deller Jr., J.H. Hansen, and J.G. Proakis. *Discete-Time Processing of Speech Signals*. IEEE Press, 2000.

[26] R.E. Kalman. Realization of covariance sequences. In *Toeplitz Memorial Conference*, Tel Aviv, Israel, 1981.

[27] J. Karlsson and A Lindquist. On degree-constrained analytic interpolation with interpolation points close to the boundary. *Provisionally accepted to IEEE Transactions of Automatic Control*, 2008.

[28] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover Publications, 1970.

[29] D. Macfarlane and K. Glover. A loop-shaping design procedure using H_∞ synthesis. *IEEE Trans. Automat. Control*, 37(6):759–769, June 1992.

[30] R. Nagamune. *Robust Control with Complexity Constraint: A Nevanlinna-Pick Interpolation Approach*. PhD thesis, Optimization and System Theory, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 2002.

[31] R. Nagamune. Closed-loop shaping based on the Nevanlinna-Pick interpolation with a degree bound. *IEEE Trans. Automat. Control*, 49:300–305, 2004.

[32] J.R. Partington. *Interpolation, Identification and Sampling*. London Mathematical Society Monographs, 1997.

[33] E. Parzen. Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2), May 1961.

[34] W. Rudin. *Real and Complex Analysis*. McGraw-Hill, 1987.

[35] A. Schuster. The periodogram of magnetic declination as obtained from the records of the greenwich observatory during the years 1871-1895. *Trans. Cambridge Philos. Soc*, 18, 1900.

[36] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[37] M. Vidyasagar and H. Kimura. Robust controllers for uncertain linear multivariable systems. *Automatica*, 22(1):85–94, January 1986.

[38] G. Vinnicombe. Frequency domain uncertainty and the graph topology. *IEEE Trans. on Automat. Control*, 38:1371–1383, 1993.

[39] G. Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms and approximate inverses. *IEEE Trans. Automat. Contr.*, 26:301–320, 1981.

The Inverse Problem of Analytic Interpolation with Degree Constraint and Weight Selection for Control Synthesis

Johan Karlsson, Tryphon Georgiou, and Anders Lindquist

Abstract

The minimizers of certain weighted entropy functionals are the solutions to an analytic interpolation problem with a degree constraint, and all solutions to this interpolation problem arise in this way by a suitable choice of weights. Selecting appropriate weights is pertinent to feedback control synthesis, where interpolants represent closed-loop transfer functions. In this paper we consider the correspondence between weights and interpolants in order to systematize feedback control synthesis with a constraint on the degree. There are two basic issues that we address: we first characterize admissible shapes of minimizers by studying the corresponding inverse problem, and then we develop effective ways of shaping minimizers via suitable choices of weights.

A.1 Introduction

The topic of this paper relates to the framework and the mathematics of modern robust control. The foundational work [37] of George Zames in the early 1980's cast the basic robust control problem as an analytic interpolation problem, where interpolation constraints ensure stability of the feedback scheme, and a norm bound guarantees performance and robustness. In this context, the analytic interpolant represents a particular transfer function of the feedback system. The work of Zames and the fact that the degree of the interpolant relates to the dimension of the closed-loop system motivated a program to investigate analytic interpolation with degree constraint (see [8, 9]). This led to an approach based on convex optimization, in which interpolants of a certain degree are obtained as minimizers of *weighted* entropy functionals. In this paper we study the correspondence between weights and such interpolants, and we develop a theory which allows for systematic shaping of interpolants to specification.

The basic issue of how the choice of weights and indices in optimization problems affects the final design is by no means new. It was R.E. Kalman [24] who, in the context of quadratic optimal control, first raised the question of what it is that characterizes optimal designs and, further, how to describe all performance criteria for which a certain design is optimal. Following Kalman's example we study here the analogous inverse problem for analytic interpolation with complexity constraint, namely the problem to decide when a particular interpolant is a minimizer of some weighted entropy functional, and if so, to determine the set of all admissible weights.

The analysis of the inverse problem leads to a new procedure for feedback control synthesis. More specifically, the quality of control depends on the frequency characteristics of the interpolant, which in turn dictates the loop shape of the feedback control system. The theory of [8, 9] provides a parametrization of all interpolants, having degree less than the number of interpolation constraints, in terms of weights in a suitable class. These admissible weights are specified by their roots. These roots coincide with the *spectral zeros* of the corresponding minimizers of the weighted entropy functionals [8]. The choice of weights for feedback control design via this procedure has been the subject of several papers (see e.g., [31, 32]). The challenge stems from the fact that the correspondence between weights and the shape of interpolants is nonlinear. One of the contributions of this paper is to develop a systematic procedure for the selection of weights based on the inverse problem.

The synthesis proceeds in two steps. We first obtain an interpolant with the required shape, but without any restriction on the degree. Then, via the inverse problem, we identify all weights for which the given interpolant is a minimizer of the corresponding entropy functional. The problem of approximating the interpolant by one of lower degree is then replaced by approximating weights in a suitable class. This approximation problem is quasi-convex and can be solved by standard methods. Hence we have replaced a non-convex problem by one that is tractable.

In Section A.2 we establish notation and review basic facts on bounded analytic interpolation and complexity-constrained interpolation. We only discuss interpolation in the unit disc $\mathbb{D} = \{z : |z| < 1\}$, but the theory applies equally well to interpolation in the half plane. In Section A.3 we consider two motivating examples in the context of robust control. In Section A.4 we provide the characterization of minimizers of weighted entropy functionals and describe the set of weights which give interpolants of a prespecified bounded degree. In Section A.5 we formulate and solve the inverse problem which is one of the key tools needed in the paper, and study the possible shapes of minimizers without any bound on their degree. In Section A.6 we develop a method for degree reduction of interpolants via the corresponding weights. In Section A.7 we highlight the steps of the synthesis procedure. Finally, in Section A.8, we revisit the motivating examples of Section A.3 and apply the procedure of Section A.7.

A.2 Background

Given complex numbers $z_0, z_1, \ldots, z_n$ in $\mathbb{D}$ which we assume to be distinct for simplicity, and given complex numbers $w_0, w_1, \ldots, w_n$, the classical Pick interpolation problem asks for a function f in the *Schur class*

$$\mathcal{S} = \{f \in H_\infty(\mathbb{D}) \,:\, \|f\|_\infty \leq 1\}$$

which satisfies the interpolation condition

$$f(z_k) = w_k, \quad k = 0, 1, \ldots, n, \tag{A.1}$$

where $H_\infty(\mathbb{D})$ (or simply H_∞) is the Hardy space of bounded analytic functions on $\mathbb{D}$. It is well-known (see, e.g., [16]) that such a function exists if and only if the Pick matrix

$$P = \left[\frac{1 - w_k \bar{w}_\ell}{1 - z_k \bar{z}_\ell}\right]_{k,\ell=0}^{n} \tag{A.2}$$

is positive semi-definite. The solution is unique if and only if P is singular, in which case f is a Blaschke product of degree equal to the rank of P. In this paper, throughout, we assume that P is positive definite and hence that there are infinitely many solutions to the Pick problem. A complete parameterization of all solutions was given by Nevanlinna (see e.g. [1]), and for this reason the subject is often referred to as Nevanlinna-Pick interpolation.

In engineering applications f usually represents the transfer function of a feedback control system or of a filter, and therefore the McMillan degree of f is of significant interest. Thus, it is natural to require that f be rational and of bounded degree. Such a constraint completely changes the nature of the underlying mathematical problem.

The Nevanlinna-Pick theory provides one particular solution that is rational and of a generic degree equal to n – the so-called *central solution*. However, it provides no insight and no help in determining any other possible solutions of the same degree. The central solution is also referred to as the *maximum-entropy solution* because it maximizes the functional

$$\int_{\mathbb{T}} \log(1 - |f|^2) dm$$

subject to (A.1), where $\mathbb{T} = \{z = e^{i\theta} \,:\, \theta \in (-\pi, \pi]\}$ is the unit circle and $dm := d\theta/2\pi$ is the (normalized) Lebesgue measure on $\mathbb{T}$. Invariably, determining extremals for such an entropy functional leads to a set of linear equations. These are often referred to as the canonical equations, and one such example is Levinson's equations in spectral analysis. For more recent developments and generalizations we refer to [4, 13, 30].

Following [9, 21], we consider the generalized entropy functional

$$\mathbb{K}_\Psi : \mathcal{S} \to \mathbb{R} \cup \infty,$$

defined by

$$\mathbb{K}_\Psi(f) = -\int_{\mathbb{T}} \Psi \log(1 - |f|^2) dm, \tag{A.3}$$

where Ψ is a non-negative log-integrable function on $\mathbb{T}$. We study how the minimizer of

$$\min\{\mathbb{K}_\Psi(f) \; : \; f \in \mathcal{S}, \; f(z_k) = w_k, \; k = 0, \ldots, n\} \tag{A.4}$$

depends on the weighting function Ψ and then determine when an interpolant f is attainable as a minimizer of (A.4) for a suitable choice of Ψ. One particularly interesting case, as we will see below, is when $\Psi = |\sigma|^2$ and σ belongs to the class of rational functions with poles at the conjugate inverses of the interpolation points.

Let ϕ be the Blaschke product

$$\phi(z) = \prod_{k=0}^{n} \frac{z_k - z}{1 - \bar{z}_k z} \tag{A.5}$$

and let $U : f(z) \mapsto z f(z)$ denote the standard shift operator on H_2. Then ϕH_2 is a shift invariant subspace, i.e. $f \in \phi H_2$ implies that $U(f) = zf \in \phi H_2$. Denote by $\mathcal{K}$ the co-invariant subspace $H_2 \ominus \phi H_2$. Then $\mathcal{K}$ is invariant under U^*, where U^* denotes the adjoint of U. Let $\mathcal{K}_0$ denote the set of outer functions in $\mathcal{K}$ that are positive at the origin. The following result is taken from [9].

Theorem A.1. *Suppose that the Pick matrix* (A.2) *is positive definite, and let σ be an arbitrary function in $\mathcal{K}_0$. Then there exists a unique pair of elements $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ such that*

(i) $f = b/a \in H^\infty$ *with* $\|f\|_\infty \leq 1$

(ii) $f(z_k) = w_k, \quad k = 0, 1, \ldots, n$, *and*

(iii) $|a|^2 - |b|^2 = |\sigma|^2$ *a.e. on* $\mathbb{T}$.

Conversely, any pair $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ satisfying (i) and (ii) determines, via (iii), a unique $\sigma \in \mathcal{K}_0$. Moreover, setting $\Psi = |\sigma|^2$, the optimization problem

$$\min \mathbb{K}_\Psi(f) \;\; \text{s.t.} \;\; f(z_k) = w_k, \; k = 0, \ldots, n$$

has a unique solution f that is precisely the unique $f \in \mathcal{S}$ satisfying conditions (i), (ii) and (iii).

We define

$$\tau(z) := \prod_{k=0}^{n} (1 - \bar{z}_k z),$$

where $\{z_k\}_{k=0}^n$ are the interpolation points. Then, it is easy to see that

$$\mathcal{K} = \left\{ \frac{p(z)}{\tau(z)} \; : \; p \in \mathrm{Pol}(n) \right\},$$

where $\mathrm{Pol}(n)$ denotes the set of polynomials of degree at most n, and

$$\mathcal{K}_0 = \left\{ \frac{p(z)}{\tau(z)} \; : \; p \in \mathrm{Pol}_+(n) \right\},$$

where $\mathrm{Pol}_+(n)$ denotes the subset of polynomials $p \in \mathrm{Pol}(n)$ such that $p(z) \neq 0$ for all $z \in \mathbb{D}$ and $p(0) > 0$.

The zeros of the function $\sigma = p/\tau \in \mathcal{K}_0$ with $p \in \mathrm{Pol}_+(n)$, as in the theorem, are the zeros of the spectral factor of the nonnegative function $1 - |f|^2$ on the unit circle. Thus, the n roots of the polynomial $z^n \bar{p}(z^{-1})$ are often referred to as *spectral zeros*. We also note that the degree of f may be less than n, which happens when a, b, and σ, have common roots.

Theorem A.1 has two parts: the first part states that interpolants of degree at most n are completely parameterized in terms of spectral zeros. More specifically, there is bijection between the pairs (b, a) such that $f = b/a$ is an interpolant of degree at most n and sets of n points in the unit disc –the roots of σ. Given such an arbitrary choice of n spectral zeros, the second part provides a convex optimization problem, the unique solution of which is precisely the corresponding interpolant.

The theorem, stated in [9], allows for considerably more general interpolation conditions than (ii). In the case where the points $\{z_0, \ldots, z_n\}$ are not necessarily distinct, condition (ii) needs to be replaced by

$$f = f_0 + \phi q \text{ with } q \in H_\infty(\mathbb{D}),$$

which encapsulates interpolation of derivatives as well. The special case where $\phi(z) = z^n$ is analogous to the so-called covariance extension problem with degree constraints, which is usually stated for Carathéodory functions rather than Schur functions. The theorem is also valid when ϕ is an arbitrary inner function. The background to the derivation of Theorem A.1 has a long history. The existence part of the parameterization was first proved in the covariance extension case in [19, 17] and in the Nevanlinna-Pick case in [18]. The uniqueness part (as well as well-posedness) in [12]. The optimization approach was initiated in [10] (also, see the extended version [11]) and further developed in, e.g., [8, 7, 20].

A.3 Motivating examples

We present two basic examples of robust control design to highlight the relevance of the theory. The first one deals with sensitivity minimization and revisits an (academic) example which was discussed in [8]. The second example addresses a more typical (and well-conditioned) design which is formulated in the context of H_∞-loopshaping and optimal robustness in the gap metric.

A.3.1 Sensitivity minimization

Consider the standard feedback system in Figure A.1 where $P(z)$ and $C(z)$ are the transfer functions of plant and controller, respectively. The stability and tracking/disturbance-rejection qualities of the feedback system are reflected in properties of the sensitivity function $S(z) = (1 - P(z)C(z))^{-1}$. It is well-known (see, for example, [38]) that the internal stability of the feedback system is equivalent to $S(z)$ being analytic outside the unit disc

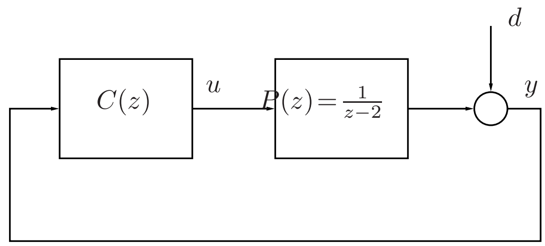

Figure A.1: Feedback system.

and satisfying the interpolation conditions

$$\begin{aligned} S(z_k) &= 1, \quad k = 1, 2, \ldots, n_z, \\ S(p_\ell) &= 0, \quad \ell = 1, 2, \ldots, n_p, \end{aligned}$$

where $z_1, \ldots, z_{n_z}$ and $p_1, \ldots, p_{n_p}$ are the zeros and poles outside the unit disk, respectively, of the plant $P(z)$. Conversely, if $S(z)$ is any stable, proper rational function which satisfies these conditions, then it can be specified as the sensitivity function of such a feedback system with the given $P(z)$ and a choice of a suitable control transfer function $C(z)$.

The maximal disturbance-amplification depends on the choice of the controller and is equal to the H_∞-norm of the sensitivity function $\|S\|_\infty$. To illustrate this, consider the simple example $P(z) := \frac{1}{z-2}$. The optimal value $\|S_{\text{opt}}\|_\infty$ in this example turns out to be equal to two (see [8]), and the magnitude of the optimal sensitivity function S_{opt} is constant across frequencies. The choice of such an optimal controller is in fact unique. However, in general, the magnitude of disturbances and the modeling uncertainty are *not* uniform across frequencies. Thus, the design ought to differentiate between frequency bands so as to achieve desired levels of performance and robustness. Therefore, we need to relax the requirement of uniformly minimal sensitivity gain to a pre-specified upper bound

$$\|S\|_\infty \leq \gamma.$$

In this case, provided $\gamma > \|S_{\text{opt}}\|_\infty$, there is large family of controllers which achieve such a specification. Naturally, the size of the family and the ability to "shape" $|S|$ increases with γ. On the other hand, the degree of the controller and the complexity of the feedback system depend on the degree of S. Thus, for a given value of γ, our goal is to not only control the shape of S, but its degree as well.

Adhering to a typical design specification for disturbance rejection, we require

$$\begin{aligned} |S(e^{i\theta})| &\leq \epsilon, \quad \text{on } (0, \theta_c), \\ |S(e^{i\theta})| &\leq \gamma, \quad \text{on } (\theta_c, \pi), \end{aligned} \tag{A.6}$$

where we take $\gamma = 2.5, \epsilon = 0.75$ and $\theta_c = 0.25$. In Figure A.2, the degree 1 sensitivity functions which satisfy $\|S\|_\infty \leq \gamma$ are depicted in Subplot 1. It is observed that the design specifications are not met by any such function. Subplot 2, in the same figure, shows a sensitivity function of degree 5 which satisfies the constraints. As expected, this shows that by relaxing the degree constraint to degree 5, we are able to find a function that satisfies the constraint. The design is accomplished with the method in Section A.6, utilizing the theory in the first part of the paper. Details will be provided in Section A.8 where the example will be revisited.

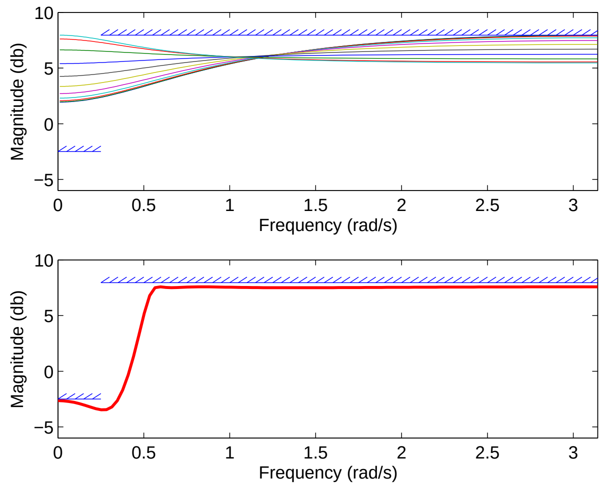

Figure A.2: Subplot 1: Possible degree one shapes. Subplot 2: A degree 5 function satisfying the constraints.

A.3.2 Frequency-dependent robustness margin

Let P_0 denote the transfer function of a single-input, single-output finite-dimensional linear system with stable coprime factorizations

$$P_0 = \frac{N_0}{M_0},$$

i.e., $M_0, N_0 \in H_\infty$, normalized to satisfy

$$M_0^* M_0 + N_0^* N_0 = 1, \text{ on } \mathbb{T}, \tag{A.7}$$

where $f(z)^* := \overline{f(\bar{z}^{-1})}$. Then, as is well-known, all stabilizing controllers for P_0 are parameterized by $q \in H_\infty$ via

$$C = \frac{U_0 + M_0 q}{V_0 + N_0 q}, \tag{A.8}$$

where $U_0, V_0 \in H_\infty$ satisfy $V_0 M_0 - U_0 N_0 = 1$, see, e.g., [14, 34]. To model perturbations of the coprime factors for frequency-dependent uncertainty, we consider plants $P = N/M$ such that

$$\left\| \begin{pmatrix} M(z) - M_0(z) \\ N(z) - N_0(z) \end{pmatrix} \right\| < \alpha |w(z)| \text{ for } z \in \mathbb{T}, \tag{A.9}$$

where $\| \cdot \|$ denotes Euclidean vector norm and w is an outer function shaping the radius. Moreover, the size of the radius is controlled by a separate scaling parameter $\alpha \in \mathbb{R}_+$. Thus, we consider the problem of robust stabilization of the ball of plants

$$\mathcal{B}(P_0, \alpha w) := \left\{ P = \frac{N}{M} \; : \; \text{(A.9) holds} \right\},$$

around the center P_0.

As shown in [35], a controller specified by q stabilizes $\mathcal{B}(P_0, \alpha w)$ provided

$$\left\| \begin{pmatrix} U_0 + M_0 q \\ V_0 + N_0 q \end{pmatrix} \alpha w \right\|_\infty \leq 1. \tag{A.10}$$

This condition can be expressed as a Nevanlinna-Pick problem. Indeed, taking advantage of the normalization of the coprime factors as in [29] (see also [22]), we define the transformation

$$Z := \begin{pmatrix} M_0^* & N_0^* \\ -N_0 & M_0 \end{pmatrix}$$

which is unitary, i.e., $ZZ^* = Z^* Z = I$. We also denote by ϕ the Blaschke product that vanishes at the conjugate inverse of the poles of M_0, N_0. Hence, $\phi M_0^*, \phi N_0^* \in H_\infty$. Then, the left hand side of (A.10) is

$$\left\| \begin{pmatrix} \phi & 0 \\ 0 & 1 \end{pmatrix} Z \begin{pmatrix} U_0 + M_0 q \\ V_0 + N_0 q \end{pmatrix} \alpha w \right\|_\infty = \left\| \begin{pmatrix} F \\ 1 \end{pmatrix} \alpha w \right\|_\infty, \tag{A.11}$$

where $F_0 = \phi M_0^* U_0 + \phi N_0^* V_0 \in H_\infty$ and

$$F = F_0 + \phi q. \tag{A.12}$$

It can be seen that the values of F at the roots of ϕ are independent of q and are specified by the plant. Moreover, as seen from (A.11), condition (A.10) holds provided $F \in H_\infty$ satisfies (A.12) and

$$\sqrt{|F|^2 + 1} \leq \frac{1}{\alpha |w|}, \text{ on } \mathbb{T}. \tag{A.13}$$

Conversely, for any $F \in H_\infty$ satisfying (A.12), there corresponds a unique parameter q and a controller C, where C stabilizes the ball of plants $\mathcal{B}(P_0, \alpha w)$ with radius

$$\alpha |w| = (|F|^2 + 1)^{-\frac{1}{2}}.$$

Furthermore, if the degree of F is small, so is the degree of the controller C. This is stated in the following proposition proved in the appendix.

Proposition A.2. *Let $F \in H_\infty$ satisfy* (A.12) *and C be the controller specified via* (A.12) *and* (A.8). *Then*

$$\deg C \leq \deg F.$$

We consider

$$\Pi_{P//C} := \begin{pmatrix} \frac{P_0}{1-P_0C} & \frac{-P_0C}{1-P_0C} \\ \frac{1}{1-P_0C} & \frac{-C}{1-P_0C} \end{pmatrix} = Z^* \begin{pmatrix} 1 & -\phi^* F \\ 0 & 0 \end{pmatrix} Z,$$

which is the matrix of transfer functions from disturbances at the input and output ports of the plant to the plant input and output. This is a rank-one matrix function (see [22]) with singular value $\sqrt{|F|^2+1}$. Thus, the shape of $|F|$ relates directly to amplification of external disturbances in the loop, and it also dictates how robust the control system is to plant uncertainty in the coprime factor (or, in the gap metric; cf. [29, 22]). In fact,

$$b_{\text{opt}}(P) := \max_{C \text{ stabilizing}} \|\Pi_{P//C}\|_\infty^{-1}$$

is precisely the optimal robustness radius for gap-ball uncertainty (see [22]) and coincides with $1/\sqrt{|F|^2+1}$ for the smallest $\|F\|_\infty$ consistent with (A.12).

The use of a frequency-dependent weight w allows shaping the loop-gain [29] as well as the performance and the robustness of the closed-loop system over different frequency bands [35, 6, 36]. By scaling α in (A.13) one can maximize the radius of $\mathcal{B}(P_0, \alpha w)$ for which a stabilizing controller exists (as in [35, 29, 22]). The maximal value α_{max} and the optimal interpolant F, consistent with (A.12) and (A.13), satisfy

$$|F|^2 = \frac{1-\alpha_{\text{max}}^2 |w|^2}{\alpha_{\text{max}}^2 |w|^2}.$$

Thus, the use of a nontrivial weight w forces the interpolant to have a nontrivial outer factor as well. This causes a corresponding increase in the degree of the closed-loop system and of the controller. Therefore, the purpose of this work is to develop techniques for reducing the degree of the control system while relaxing design requirements in a controlled fashion. In Section A.8 we revisit the issue of frequency-dependent robustness margin in order to demonstrate the application of the framework.

A.4 Characterization of $\mathbb{K}_\Psi$-minimizers

Theorem A.1 provides a (complete) parametrization of Nevanlinna-Pick interpolants of degree $\leq n$. It states that such interpolants are in correspondence with $\Psi = |\sigma|^2$ for $\sigma \in \mathcal{K}_0$. Furthermore, it states that such interpolants originate as minimizers of the generalized entropy integral $\mathbb{K}_\Psi$ specified by such a weight Ψ. In this paper we are also interested in interpolants of higher degree. Thus, we are led to consider $\mathbb{K}_\Psi$-entropy minimizers for more general choices of Ψ. Indeed, the entropy functional $\mathbb{K}_\Psi$ can be defined for arbitrary nonnegative functions Ψ and since the minimizer relates algebraically to Ψ, choices of a rational Ψ generate minimizers of a suitable degree. Thus, the following theorem is a generalization of Theorem A.1 which allows for the use of general weights Ψ. This is one of our main results.

Theorem A.3. *Suppose that the Pick matrix* (A.2) *is positive definite and* Ψ *is a log-integrable nonnegative function on the unit circle. A function* f *is a minimizer of* (A.4) *if and only if the following three conditions hold:*

(i) $f(z_k) = w_k$ *for* $k = 0, \dots, n$*,*

(ii) $f = \frac{b}{a} \in \mathcal{S}$ *where* $b \in \mathcal{K}$ *and* a *is outer,*

(iii) $\Psi = |a|^2 - |b|^2$*.*

Any such minimizer is necessarily unique.

Proof. See the appendix. □

As seen from Theorem A.1, any choice of $\Psi = |\sigma|^2$ with $\sigma \in \mathcal{K}_0$ gives rise to $a, b \in \mathcal{K}$, and hence to an interpolant f with a degree bounded by n. Even if Ψ is rational of an arbitrarily high degree or is irrational, b still belongs to $\mathcal{K}$. In fact, the additional "complexity" is absorbed in a. Naturally, in such a case, the interpolant f will also be rational of a high degree or irrational, respectively. This observation allows us to characterize all minimizers of $\mathbb{K}_\Psi$ of degree at most $n + m$ for any given $m \in \mathbb{N}_+$. More specifically, let

$$\begin{aligned} \mathcal{K}_m &:= \left\{\sigma = \sigma_0 p \,:\, \sigma_0 \in \mathcal{K}_0,\, \deg p \leq m,\, p \text{ outer},\, p(0) > 0\right\} \\ &= \left\{\frac{q}{\tau\pi} \,:\, \pi \in \text{Pol}_+(m),\, q \in \text{Pol}_+(n+m)\right\} \end{aligned} \tag{A.14}$$

The following statement gives the sought characterization.

Proposition A.4. *Let* $\Psi = |\sigma|^2$ *with* $\sigma \in \mathcal{K}_m$*. Then the minimizing function* f *in* (A.4) *satisfies*

(i) $f(z_k) = w_k$ *for* $k = 0, \dots, n$*,*

(ii) f *has at most* n *zeros in* $\mathbb{D}$*,*

(iii) the degree of f *is at most* $n + m$*.*

Conversely, for any $f \in \mathcal{S}$ *which satisfies (i), (ii), and (iii), there exists a corresponding choice of* $\sigma \in \mathcal{K}_m$ *so that* f *is the minimizer of* (A.4)*.*

Proof. By Theorem A.3, the minimizer f satisfies (ii), where

$$|a|^2 = |\sigma|^2 + |b|^2,\ \sigma \in \mathcal{K}_m, b \in \mathcal{K}.$$

Then, in view of (A.14), $a \in \mathcal{K}_m$, and hence $\deg f \leq n + m$. Conversely, suppose $f = (\beta\pi)/\alpha$, where $\beta \in \text{Pol}(n)$, $\pi \in \text{Pol}_+(m)$ and $\alpha \in \text{Pol}_+(n+m)$, then $f = b/a$, where $b \in \mathcal{K}$ and $a \in \mathcal{K}_m$. Then

$$\Psi = |a|^2 - |b|^2 = |\sigma|^2,$$

where $\sigma \in \mathcal{K}_m$. If, in addition, f satisfies the interpolation conditions in (i), then, by Theorem A.3, f is the minimizer of (A.4), as claimed. □

Corollary A.5. *If $f \in \mathcal{S}$ is a minimizer of $\mathbb{K}_\Psi$ for some choice of a log-integrable non-negative function Ψ, then f has at most n zeros in $\mathbb{D}$.*

Proof. The corollary follows directly from condition (ii) in Theorem A.3, since $f = b/a$ with a outer and $b \in \mathcal{K}$. □

This corollary underscores the significance of Theorem A.3 for understanding the structure of minimizers.

A.5 The inverse problem of analytic interpolation with degree constraint

We begin by considering the inverse problem of analytic interpolation with a degree constraint, and then explain how this can be used to shape interpolants to specifications.

A.5.1 The inverse problem

It turns out that the number of roots in $\mathbb{D}$ determine whether an interpolant f is a minimizers of $\mathbb{K}_\Psi$ for some choice of Ψ. Furthermore, it is possible to characterize all such Ψ. This is stated below.

Proposition A.6. *Any function $f \in \mathcal{S}$ that satisfies*

(i) $f(z_k) = w_k$ *for* $k = 0, \ldots, n$,

(ii) f *has at most* n *zeros in* $\mathbb{D}$,

(iii) $\log(1 - |f|^2) \in L_1(\mathbb{T})$,

is the unique minimizer of (A.4) *with*

$$\Psi = (|f|^{-2} - 1)|b|^2$$

for any $b \in \mathcal{K}$ chosen so that bf^{-1} is outer. Conversely, a (nonzero) function f having more than n zeros in $\mathbb{D}$ cannot arise as the minimizer of (A.4) *for any choice of Ψ.*

Proof. Suppose that $f \in \mathcal{S}$ satisfies the interpolation constraint (i). Then, by Theorem A.3, f is the minimizer of (A.4) if and only if $f = b/a$, where $b \in \mathcal{K}$, a is outer, and $\Psi = |a|^2 - |b|^2$, which in turn holds if and only if bf^{-1} is outer, $b \in \mathcal{K}$, and $\Psi = (|f|^{-2} - 1)|b|^2$. This condition cannot hold if f has more than n zeros in $\mathbb{D}$. In fact, for bf^{-1} to be outer, the zeros of f in $\mathbb{D}$ must be canceled by zeros of b. However, $b \in \mathcal{K}$ can have at most n zeros. □

The choice of $b \in \mathcal{K}$ in Theorem A.6 is not unique, in general. The selection of b must prevent bf^{-1} from having poles in $\mathbb{D}$, and hence any zero of f must also be a zero of b. If f has more than n zeros in $\mathbb{D}$, there is no such b, whereas if f has exactly n zeros in $\mathbb{D}$, then b is uniquely defined. In all other cases, when f has $n_f < n$ zeros in $\mathbb{D}$, the family of possible choices of b, and hence the family of possible weights

$$\{\Psi \,:\, \Psi = (|f|^{-2} - 1)|b|^2, b \in \mathcal{K},\, bf^{-1} \text{ outer}\}$$

has dimension $n - n_f$. The design freedom offered by this nonuniqueness will be exploited in Section A.7 for finding a weight corresponding to f that is close to a low-degree weight.

Remark A.1. Since the more general frequency-dependent constraint

$$|f(z)| \leq |w(z)| \text{ for } z \in \mathbb{T},$$

with w an outer weighting function, is often quite natural, [21] introduced the (doubly "weighted") entropy functional $-\int_{\mathbb{T}} \Psi \log(1 - |w^{-1}f|^2)dm$. Thus, compared to (A.3), w can be introduced as an extra design-parameter. Proposition A.6 may be modified accordingly by changing condition (iii) to $\log(1 - |w^{-1}f|^2) \in L_1(\mathbb{T})$. Then f is the unique minimizer of

$$\min \mathbb{K}_{\Psi}(w^{-1}f) \text{ s. t. } f(z_k) = w_k,\ k = 0, \ldots, n, \tag{A.15}$$

for the weight

$$\Psi = \left(\left|w^{-1}f\right|^{-2} - 1\right)|b|^2$$

with $b \in \mathcal{K}$ chosen so that bf^{-1} is outer. The price to pay is that, in general, the degree of the interpolant f increases by the degree of w.

A.5.2 Shaping the interpolant based on the inverse problem

Suppose that g is a given function on $\mathbb{T}$. We address the following two questions that pertain to admissible shapes of analytic interpolants:

a) Does there exist an $f \in \mathcal{S}$ which satisfies (A.1) and

$$|f(e^{i\theta})| \leq |g(e^{i\theta})|, \quad \theta \in (-\pi, \pi]? \tag{A.16}$$

b) Does there exist a Ψ such that the corresponding (unique) minimizer f of (A.4) satisfies

$$|f(e^{i\theta})| = |g(e^{i\theta})|, \quad \theta \in (-\pi, \pi]? \tag{A.17}$$

Without loss of generality we may assume that g is an outer function and that, moreover, $g \in \mathcal{S}$. We note that the first question is equivalent to finding an f in the set

$$\{f \in \mathcal{S} : \|fg^{-1}\|_\infty \leq 1, f(z_k) = w_k, k = 0, \ldots, n\}, \tag{A.18}$$

and this set is nonempty if and only if the associated Pick matrix

$$\text{Pick}(g) := \left[\frac{1 - w_k g(z_k)^{-1}\overline{w_l g(z_l)^{-1}}}{1 - z_k\overline{z_l}}\right]_{k,l=0}^{n} \tag{A.19}$$

is positive semi-definite. This answers the first question.

If Pick(g) is positive definite, there exist interpolants $\hat{f}$ such that $|\hat{f}(z)| < |g(z)|$ for all $z \in \mathbb{T}$ and the design specifications (A.16) may be satisfied with strict inequality. Therefore, any minimizing interpolant f cannot satisfy (A.17), since $|\hat{f}(z)| < |f(z)|$ for all $z \in \mathbb{T}$ implies that $\mathbb{K}_\Psi(\hat{f}) < \mathbb{K}_\Psi(f)$, which contradicts the claim that f is the minimizer. Therefore in order for (A.4) to have a solution satisfying (A.17), Pick(g) must be positive semidefinite and singular. In this case, the set (A.18) is a singleton. This provides an answer to the second question, which we summarize next.

Proposition A.7. *Let $g \in \mathcal{S}$ be outer and such that $\log(1 - |g|^2) \in L_1$. Then there exists a pair (Ψ, f) of functions on $\mathbb{T}$ such that*

(i) $\log \Psi \in L_1$,

(ii) f is the solution of (A.4), and

(iii) $|f| = |g|$ on $\mathbb{T}$

if and only if Pick(g) is positive semidefinite and singular. Furthermore, f is uniquely determined.

Proof. Sufficiency is shown in the text leading to the proposition, so it remains to prove necessity. Since the matrix in (A.19) is nonnegative definite and singular, there is a unique f satisfying $\|fg^{-1}\| \leq 1$ and (A.1). Then $f = g\varphi$ where φ is inner and of degree $\leq n$. Since f is rational with at most n zeros in $\mathbb{D}$, by Proposition A.6 there exists a functions Ψ such that f is the minimizer of (A.4). □

A.6 Continuity properties of the map from weights to minimizers

Assume that f is the minimizer of the entropy functional, as in (A.4), for a suitable weight selected without regard to the degree. We begin by studying the properties of the nonlinear transformation

$$\varphi \,:\, \Psi \mapsto f \tag{A.20}$$

which maps a space of weights

$$\Psi \in \mathcal{M} := \{\Psi \mid \log \Psi \in L_\infty(\mathbb{T})\}$$

to the corresponding minimizers f of (A.4). We define the metric on $\mathcal{M}$ as $d(\Psi, \Psi_r) := \|\log(\Psi) - \log(\Psi_r)\|_\infty$, and, as we shall see, φ is continuous when the range is taken to be H_2, but not with range H_∞. On the other hand, the map $\Psi \mapsto |f| \in L_\infty$ is again continuous.

Remark A.2. Here we only study the case where $\Psi \in \mathcal{M}$. However, it is easy to show that these continuity properties also holds for arbitrary L_∞ perturbations of $\log \Psi$ when Ψ is non-negative and log-integrable.

A.6.1 Continuity of the map φ

In this subsection we establish the continuity of the map $\varphi : \mathcal{M} \to H_2$. Moreover, with a counterexample, we show that the corresponding map $\varphi : \mathcal{M} \to H_\infty$, with the range endowed with the topology of H_∞, is not continuous.

Lemma A.8. *Let Ψ and Ψ_r be nonnegative log-integrable functions on $\mathbb{T}$ that satisfy*

$$\| \log(\Psi) - \log(\Psi_r) \|_\infty = \epsilon, \tag{A.21}$$

and set $f := \varphi(\Psi)$ and $f_r := \varphi(\Psi_r)$. Then

$$\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f - f_r|^2}{8}\right) dm \leq (e^{2\epsilon} - 1)\mathbb{K}_\Psi(f). \tag{A.22}$$

Proof. From (A.21), it follows that

$$e^{-\epsilon}\Psi(z) \leq \Psi_r(z) \leq e^{\epsilon}\Psi(z), \text{ for } z \in \mathbb{T}$$

and hence

$$\mathbb{K}_\Psi(f_r) \leq e^{\epsilon}\,\mathbb{K}_{\Psi_r}(f_r) \leq e^{\epsilon}\,\mathbb{K}_{\Psi_r}(f) \leq e^{2\epsilon}\,\mathbb{K}_\Psi(f).$$

From [27] we have that

$$\frac{1}{2}(\mathbb{K}_\Psi(f) + \mathbb{K}_\Psi(f_r)) \geq \mathbb{K}_\Psi\left(\frac{f + f_r}{2}\right) + \frac{1}{2}\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f - f_r|^2}{8}\right) dm.$$

Then, since f is the minimizer of $\mathbb{K}_\Psi$, we have

$$\mathbb{K}_\Psi(f) \leq \mathbb{K}_\Psi\left(\frac{f + f_r}{2}\right),$$

and consequently

$$\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f - f_r|^2}{8}\right) dm \leq \mathbb{K}_\Psi(f_r) - \mathbb{K}_\Psi(f) \leq (e^{2\epsilon} - 1)\mathbb{K}_\Psi(f),$$

as claimed. □

Lemma A.8 provides a bound on a weighted H_2 norm of the error $f - f_r$, as stated in the next corollary.

Corollary A.9. *Let f, f_r and Ψ, Ψ_r be as in Lemma A.8 and let σ be the outer spectral factor of Ψ; i.e., σ is outer and $|\sigma|^2 = \Psi$ on $\mathbb{T}$. Then*

$$\|\sigma(f - f_r)\|_2^2 \leq 10(e^{2\epsilon} - 1)\mathbb{K}_\Psi(f). \tag{A.23}$$

Proof. Since $\log(1 + t) \geq 0.81t$ for $t \in [0, \frac{1}{2}]$, we have that

$$\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f - f_r|^2}{8}\right) dm \geq \int_{\mathbb{T}} \Psi \frac{|f - f_r|^2}{10} dm.$$

Then (A.23) follows from Lemma A.8. □

The continuity of φ is then a direct consequence of this.

Proposition A.10. *The map* φ *in* (A.20) *with range* H_2*, is continuous.*

The mapping φ from $\Psi \in \mathcal{M}$ to $f \in H_\infty$ is not continuous. This can be seen from the following counterexample. Consider the interpolation problem with one interpolation condition: $f(0) = 1/2$. Given a weight Ψ, according to Theorem A.3, the minimizer f satisfies

$$|f|^2 = \frac{1}{1 + \frac{\Psi}{|b|^2}}, \tag{A.24}$$

where $b \in \mathcal{K}$, which implies that b is a constant. Since there is only one interpolation point, and thus $n = 0$, the minimizer f cannot have any zeros (Corollary A.5). Hence f is the outer factor of (A.24), where b is a constant chosen so that the interpolation constraint is satisfied. If $\Psi \equiv 1$, then $f \equiv 1/2$. Define

$$\Psi_\epsilon(e^{i\theta}) = \begin{cases} 1 & \text{for} \quad \theta \in [-\frac{\pi}{2}, \frac{\pi}{2}] \\ 1 + \epsilon & \text{for} \quad \theta \in (\frac{\pi}{2}, \frac{3\pi}{2}) \end{cases}$$

and let $f_\epsilon = \frac{b_\epsilon}{a_\epsilon} = \varphi(\Psi_\epsilon)$, with $b_\epsilon \in \mathcal{K}$ and a_ϵ outer, as in Theorem A.3. By Lemma A.17, $b_\epsilon \to b$, but since $|f_\epsilon|^2$ is a discontinuous at $\theta = \frac{\pi}{2}$ for $\epsilon > 0$, the spectral factor f_ϵ does not converge to $f \equiv 1/2$ in H_∞ as $\epsilon \to 0$. This is since the phase of the spectral factor will differ by an arbitrarily large amount [2, page 147] (c.f. example in [2] page 145-158).

A.6.2 Continuity of the map defining the shape of the interpolant

From an engineering viewpoint, ∞-norm bounds on the approximation error are important in order to guarantee performance and robustness. Here we will show that a small approximation error on the weight Ψ will in fact correspond to a small L_∞ error in the shape of the interpolant. To this end, consider the mapping

$$\psi \;:\; \Psi \mapsto |f| \in L_\infty, \tag{A.25}$$

which maps a choice of weight $\Psi \in \mathcal{M}$ to the magnitude $|f|$ of the corresponding minimizer f of (A.4). As seen from the previous example, the lack of H_∞ continuity of interpolants is due to the fact that spectral factorization is not continuous. This problem does not ocur with ψ, which is continuous.

Proposition A.11. *The map* ψ *in* (A.25) *is continuous.*

Proof. Let $\Psi \in \mathcal{M}$ and $f = \frac{b}{a} = \varphi(\Psi)$, and let $\Psi_k \in \mathcal{M}$ and $f = \frac{b_k}{a_k} = \varphi(\Psi_k)$ for $k = 1, 2, \ldots$. Assume further that

$$\|\Psi - \Psi_k\| \to 0 \text{ as } k \to \infty.$$

Then, by Lemma A.17 in the appendix, $b_k \to b$. The proposition now follows since

$$|f_k|^2 = \frac{1}{1 + \left|\frac{\Psi_k}{b_k}\right|^2} \to \frac{1}{1 + \left|\frac{\Psi}{b}\right|^2} = |f|^2$$

uniformly. $\square$

A.7 Approximation of interpolants

The continuity properties described in the previous section suggests a new approach for approximating interpolants that exploits the correspondence between minimizers and weights. Given an interpolant f we would like to find a degree-r approximating interpolant f_r of f, where $r \geq n$. From the inverse problem there is a set $\varphi^{-1}(f)$ of admissible weights Ψ for which a given f is the minimizer of (A.4). Our first task is to find a pair (Ψ, Ψ_r) for which $\Psi \in \varphi^{-1}(f)$ and $\Psi_r = |\sigma_r|^2$, with $\sigma_r \in \mathcal{K}_{r-n}$, so that their logarithmic distance is minimal. That is, we solve the following optimization problem

$$\min \Big\{ \| \log(\Psi) - \log(\Psi_r) \|_\infty \;:\; \Psi \in \varphi^{-1}(f) \text{ and } \Psi_r = |\sigma_r|^2 \text{ with } \sigma_r \in \mathcal{K}_{r-n} \Big\}.$$

This optimization problem may be reformulated as a quasi-convex optimization and solved efficiently. By Proposition A.4 the degree of the interpolant $f_r = \varphi(\Psi_r)$ is bounded by r, and by Proposition A.10, a bound on the approximation error $f - f_r$ is obtained. In fact, based on the quality of approximation obtained via quasi-convex optimization, explicit a bound on the error can be obtained from Lemma A.8 and Corollary A.9.

By Proposition A.4, the function φ, defined in (A.20), maps $\mathcal{K}_{r-n}$ into the set of interpolants of degree at most r. Thus, the basic idea is to replace the hard nonconvex problem of approximating f by another interpolating function f_r of degree at most r, by the simpler quasi-convex problem to approximate a $\Psi \in \varphi^{-1}(f)$ by a $\Psi_r = |\sigma_r|^2$ with $\sigma_r \in \mathcal{K}_{r-n}$.

The theory presented so far suggests a computational procedure in several steps, which we now summarize. In general, the required bound on the norm of the interpolant may differ from one, and therefore we consider the more general problem to find a function F, of a desired shape, which satisfies $\|F\|_\infty \leq \gamma$ and the interpolation conditions

$$F(z_k) = W_k \text{ for } k = 0, 1 \ldots, n. \tag{A.26}$$

Step 1. Find an interpolant having the desired shape, but without restricting its degree. To this end, we begin with a family of functions $\{g_\alpha\}$ having desired shape, and we select one function g in this class for which the Pick condition in Proposition A.7 is satisfied. Then, by Proposition A.7, there is a Ψ such that $f := \varphi(\Psi)$ satisfies $|f(z)| = |g(z)|$ for all $z \in \mathbb{T}$.

Typically we choose this family of functions in such a way that $|g_\alpha|$ is monotonically decreasing in α. In the first of our motivating examples we take

$$g_\alpha = \frac{w}{\alpha},$$

which leads to a standard H_∞ optimization problem. In the second example we take

$$|g_\alpha|^2 = \frac{1 - \alpha^2 |w|^2}{\alpha^2 |w|^2}$$

which leads to a typical 2-block problem.

We then seek a solution to the optimization problem

$$\begin{aligned}&\max \alpha\\ &\text{ subject to}\\ &|F| \leq |g_\alpha| \text{ and } F(z_k) = W_k,\ k = 0, 1 \ldots, n.\end{aligned}$$

The optimum is attained when Pick(g_α) is positive semidefinite and singular (Proposition A.7). Here the bound γ must satisfy

$$\|F\|_\infty \leq \gamma \text{ and } \log(\gamma^2 - |F|^2) \in L_1,$$

or else the bound γ needs to be relaxed. Then the normalized interpolant

$$f := \frac{1}{\gamma} F$$

has the same shape as F and satisfies the interpolation conditions

$$f(z_k) = w_k := \frac{1}{\gamma} W_k,\ k = 0, 1, \ldots, n,$$

as well as the log-integrability condition of $1 - |f|^2$. Hence we have constructed an interpolant with the required shape, but which in general does not satisfy the desired degree constraint.

Step 2. For some $r \geq n$, find an approximation f_r of f of degree at most r which satisfies the same interpolation conditions. To this end, find functions Ψ and Ψ_r that solve the optimization problem

$$\begin{aligned}&\min \|\log(\Psi) - \log(\Psi_r)\|_\infty\\ &\text{ subject to}\\ &\Psi \in \varphi^{-1}(f) \text{ and } \Psi_r = |\sigma_r|^2 \text{ with } \sigma_r \in \mathcal{K}_{r-n},\end{aligned}$$

where

$$\varphi^{-1}(f) = \{\Psi \,:\, \Psi = (|f|^{-2} - 1)|b|^2 \mid b \in \mathcal{K},\ bf^{-1} \text{ outer}\}.$$

This is a quasi-convex optimization problem. In fact, $\|\log(\Psi) - \log(\Psi_r)\| \leq \epsilon$ if and only if

$$e^{-\epsilon} \leq \frac{\Psi_r(z)}{\Psi(z)} \leq e^{\epsilon} \text{ for all } z \in \mathbb{T}. \tag{A.27}$$

The constraints (A.27) define an infinite set of linear constraints on the pseudo-polynomials representing the nominator and denominator, respectively, of Ψ_r/Ψ. Since the sublevel set of nominators and denominators solving (A.27) is convex for each $\epsilon > 0$, the problem is quasiconvex. The reader is referred to [28] for a detailed description how to solve this problem.

Step 3. Next we solve the optimization problem

$$\min\{\mathbb{K}_{\Psi_r}(f_r) \,:\, f_r \in \mathcal{S},\ f_r(z_k) = w_k,\ k = 0, \ldots, n\}$$

for the unique solution f_r. In Step 2 we have determined the weight Ψ_r as an approximation of Ψ, and therefore f_r will also be an approximation of to f, for which the bounds (A.22) and (A.23) hold. Furthermore, since $\Psi_r = |\sigma_r|^2$ with $\sigma_r \in \mathcal{K}_{r-n}$, the degree of f_r is bounded by r (Proposition A.4). Finally, we renormalize the interpolant

$$F_r := \gamma f_r$$

to obtain the approximant which solves the original interpolation problem.

A.8 Examples Revisited

We now return to the two examples from Section A.3. In both examples, the underlying mathematical problem is an analytic interpolation problem where a desired shape is sought for the interpolant. These are addressed using the procedure outlined in Section A.7.

A.8.1 Sensitivity minimization (continued)

In this example, we consider the sensitivity function $S = (1 - PC)^{-1}$ of the feedback system with plant

$$P(z) = \frac{1}{z-2}.$$

Since P has one unstable pole at 2 and an unstable zero at ∞, we require that the sensitivity function satisfies

$$S(\infty) = 1 \text{ and } S(2) = 0.$$

We further require that the specifications (A.6) are met. The relaxed bound on the infinity-norm of S is $\|S\|_\infty \le \gamma := \frac{5}{2}$, and we therefore define the function

$$f(z) = \gamma^{-1} S(z^{-1})$$

which is normalized so that f satisfies $\|f\|_\infty \le 1$ and is analytic in $\mathbb{D}$. The constraints on S can be directly translated into constraints on f:

$$f(0) = 0.4, \text{ and } f(0.5) = 0,$$

and

$$\begin{aligned} |f(e^{i\theta})| &\le \epsilon\gamma^{-1}, && \text{on } (0, \theta_c), \\ |f(e^{i\theta})| &\le 1, && \text{on } (\theta_c, \pi), \end{aligned}$$

where, as before, $\gamma = 2.5, \epsilon = 0.75$ and $\theta_c = 0.25$. We begin with a particular interpolant f_{ideal} that meets the criteria without regard to any constraint on the degree, shown in Figure A.3, which we then approximate using the theory in Section A.6.

We determine f_{ideal} as follows. We first define an outer function w with the property that $\log|w(e^{i\theta})|$ is piecewise linear in $\theta \in (0, \pi)$ and passes through the points

$$(0, \epsilon\gamma^{-1}),\ (\theta_c + 0.1, \epsilon\gamma^{-1}),\ (\theta_c + 0.3, 1),\ \text{and } (\pi, 1).$$

This is our desired shape. Then we set $g = \frac{w}{\alpha}$ and scale $\alpha > 0$ so that there exists a minimizer of (A.4) which satisfies $|f| = |g|$ on $\mathbb{T}$. By Proposition A.7, α is specified by the requirement that Pick(g) is positive semidefinite and singular. In this case, $\alpha = 1.0498 > 1$, and hence $|f_{\text{ideal}}| = |g|$ is consistent with the requirement $\|f_{\text{ideal}}\|_\infty \leq 1$. It is clear that neither g nor f_{ideal} is a rational function, but, unlike g, f_{ideal} is an interpolant.

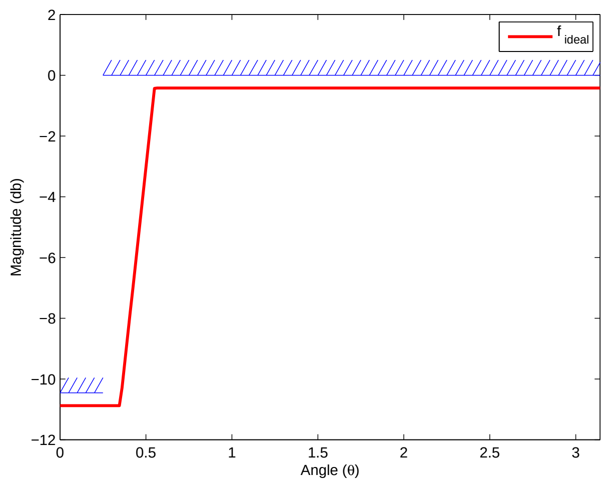

Figure A.3: Ideal function: $f_{\text{ideal}} = \gamma^{-1} S^*_{\text{ideal}}$

Next we approximate f_{ideal} by an interpolant of small degree. We first characterize the inverse image of f_{ideal} under the map (A.20), which according to Proposition A.6 is given by

$$\varphi^{-1}(f_{\text{ideal}}) = \{\Psi \,:\, \Psi = (|f_{\text{ideal}}|^{-2} - 1)|b|^2,\, b \in \mathcal{K},\, b f_{\text{ideal}}^{-1} \text{ outer}\}.$$

In the present case, $|b|^2$ is a positive constant. Hence, $\varphi^{-1}(f_{\text{ideal}})$ contains a single element, modulo scaling, and we choose $\Psi_{\text{ideal}} = (|f_{\text{ideal}}|^{-2} - 1)$. As described in Section A.5.2 we let f_r be the approximant of degree less or equal to r obtained from

$$f_r = \arg\min \left\{ \mathbb{K}_{|\sigma_r|^2}(f_r) \,\middle|\, f_r \in \mathcal{S} \text{ and (A.1) holds} \right\},$$

where σ_r is the solution of the quasi-convex optimization

$$\min \left\{ \left\| \log(|\sigma_r|^2) - \log(\Psi_{\text{ideal}}) \right\|_\infty \,\middle|\, \text{for } \sigma_r \in \mathcal{K}_{r-n} \right\}.$$

Finally, by scaling $S_r(z) := \gamma f_r(z^{-1})$ and $S_{\text{ideal}}(z) := \gamma f_{\text{ideal}}(z^{-1})$, we obtain admissible sensitivity functions.

We compute S_r for $r = 1, 3, 5$ and display the magnitudes of S_1, S_3 and S_5 in Figure A.4. Neither the degree-one nor the degree-three approximant of S_{ideal} satisfies the design specifications, wheras S_5 does. It is interesting to note that even though f_{ideal} is infinite-dimensional it is possible to find satisfactory low-dimensional approximants.

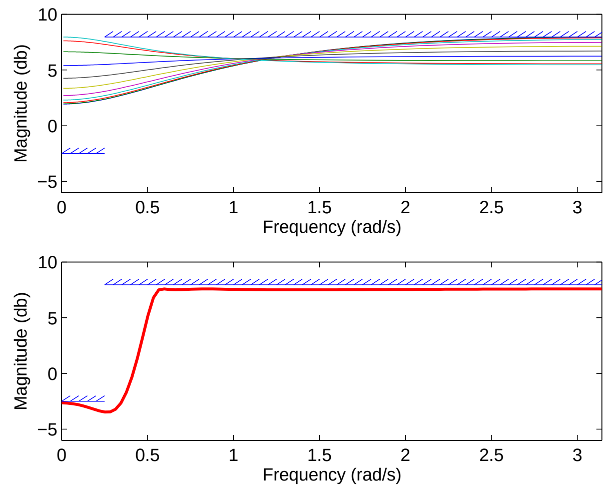

Figure A.4: Approximations of degree $1, 3$, and 5

It is also interesting to note that when the bound $\|S\|_\infty \leq \gamma$ is removed and only stability is required, as in [28], the approximation is better in the low-frequency band $(0, \theta_c + 0.1)$, but worse in the high-frequency band $(\theta_c + 0.3, \pi)$. It seems as if approximation with a bound puts more emphasis on the region where the interpolant is close to the bound (i.e. $|S| \approx \gamma$) at the expense of the region where $|S_{\text{ideal}}| \ll \gamma$.

A.8.2 Frequency-dependent robustness margin (continued)

We consider a continuous-time plant having one integrator, a slow unstable pole, and a time-lag, modeled via $\frac{s-1}{s+1}\frac{1}{s(s-0.1)}$. We base our design on its discrete-time counterpart

$$P_0(z) = \frac{0.08772z^3 - 0.08772z^2 - 0.4386z - 0.2632}{z^3 - 2.439z^2 + 1.807z - 0.3684}$$

obtained via the Möbius transform

$$s \to z = \frac{2+s}{2-s},$$

and restrict our analysis to the discrete-time domain.

The design objective is encapsulated in the choice of a weight w, chosen as in Section A.3 to increase robustness to high-frequency modeling uncertainty. The selected "nominal weight" w is shown in Figure A.5. The maximal scaling parameter α_{max} can be readily computed by first calculating the outer factor g_α of

$$|F|^2 = \frac{1-\alpha^2|w|^2}{\alpha^2|w|^2} = |g_\alpha|^2 \text{ on } \mathbb{T},$$

and then, finding the maximal value α_{max} for which the Pick matrix Pick(g_α) is positive semidefinite (cf. Proposition A.7). The Pick matrix Pick(g_α) is defined in (A.19) and requires the interpolation data that can be obtained by evaluating F_0 at the roots of ϕ in (A.12). Denote by F_{ideal} the unique interpolant which satisfies $|F_{\text{ideal}}| = |g_{\alpha_{\text{max}}}|$, and denote the corresponding controller by C_{ideal}. Since w is not rational, neither are F_{ideal} and C_{ideal}. Next we describe how to approximate F_{ideal} with an admissible interpolant of low degree. Using the corresponding controller leads to closed-loop transfer functions of low degree.

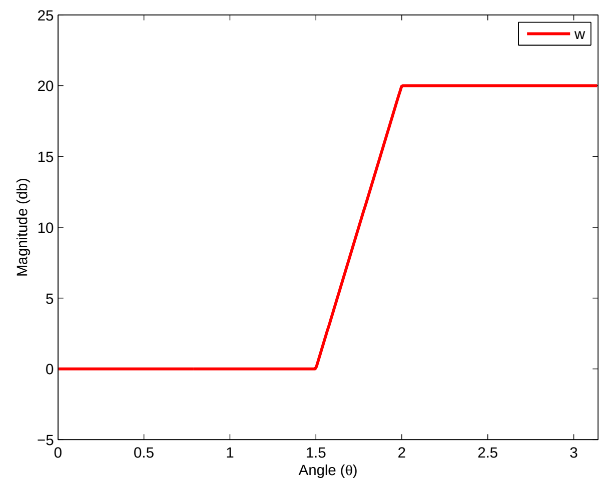

Figure A.5: The frequency-dependent robustness shape w

The uniform robustness margin corresponding to the controller C_{ideal} is determined by the value of $\|F_{\text{ideal}}\|_\infty$ via (A.13). For the above choices $\|F_{\text{ideal}}\|_\infty = 10.87$. In order to achieve the desired characteristic for the frequency-dependent robustness margin, we relax the H_∞ bound on F. For the particular example, it is deemed appropriate to allow $\|F\|_\infty \leq \gamma = 20$. We normalize F by defining $f = \frac{1}{\gamma}F$, which is then required to satisfy

$$f(z) = \frac{1}{\gamma}F_0(z) \text{ whenever } \phi(z) = 0, \tag{A.28}$$

in view of the interpolation condition (A.12). Then we follow the steps of Section A.7 to obtain approximants to $f_{\text{ideal}} = \frac{1}{\gamma} F_{\text{ideal}}$.

For any given $r \geq n$ we determine a degree-r approximant f_r of f_{ideal} as follows. We first compute a minimizer σ_r of the quasi-convex optimization problem to find a $\sigma_r \in \mathcal{K}_{r-n}$ and a $\Psi \in \varphi^{-1}(f_{\text{ideal}})$ which minimize

$$\left\| \log(|\sigma_r|^2) - \log(\Psi) \right\|_\infty .$$

Next we determine f_r as the minimizer of the convex optimization problem

$$f_r = \arg\min\{\mathbb{K}_{|\sigma_r|^2}(f_r) \big| f \in \mathcal{S} \text{ and (A.28)}\}.$$

A corresponding controller C_r can now be determined via q from (A.12) and (A.8).

The uniform robustness radius for gap-metric uncertainty is maximal for an optimal choice of the controller C_{opt} and equals $b_{\text{opt}}(P)$ ([35, 22]. This is the inverse of the H_∞-norm of the "parallel projection" operator $\Pi_{P//C_{\text{opt}}}$, and this value is shown in Figure A.6 with a dash-dotted line. On the other hand, the inverse of the maximal singular value of $\Pi_{P//C_{\text{ideal}}}$, plotted as a function of frequency with solid line, represents a frequency-dependent robustness radius [35]. Both are now compared with a degree-four approximant

$$C_4 = \frac{0.876z^4 - 0.190z^3 - 0.0669z^2 - 0.460z + 0.157}{z^4 + 0.1205z^3 + 1.389z^2 + 0.07538z + 0.2214},$$

and it is seen that there is substantial improvement of robustness as compared to $b_{\text{opt}}(P)$ in the high-frequency range. Figure A.7 compares the gains of C_4 and C_{opt}. Similarly, Figure A.8 and Figure A.9 compare the loop-gains and the Nyquist plots, respectively, for the two cases. It is seen that some form of phase compensation is effected by C_4 around 1.6 rad/sec, as compared to C_{opt} so as to gain the sought advantage. Figure A.10 compares the gains of the four entries of the closed-loop transfer matrix $\Pi_{P//C}$. The improvement in the sensitivity function at middle range becomes evident.

A.9 Concluding remarks

The formulation of feedback control synthesis as an analytic interpolation problem has been at the heart of modern developments in robust control. Yet, many of the standard approaches often lead to designs of a large degree, due to degree inflation when introducing and absorbing "weights" into the controller. At various stages, alternative methodologies for dealing with control design under structural and dimensionality constraints were developed by several authors based primarily on suitable approximations and a linear matrix inequality formalism (see [15], [33], [23], [3]). In particular, a comparison between the viewpoint in Gahinet and Apkarian [15] and Skelton, Iwasaki, and Grigoriadis [33] and the viewpoint advocated in our work is provided in [21].

Our approach builds on the original H_∞-formulation of control synthesis as an analytic interpolation problem and on the recently discovered fact that, in contrast to H_∞-minimization, dimensionality and performance are inherited by the weighted-entropy minimization. In this setting, "weights" provide the means of shaping interpolants in a manner

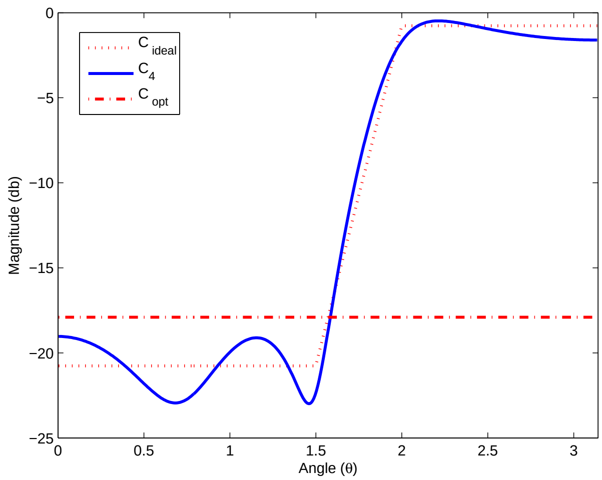

Figure A.6: The robustness radius obtained for the controllers C_{ideal}, C_4, and C_{opt}.

akin to H^∞ design. Thus, the advantage of the new methodology which involves entropy functionals stems from the fact that selection of weights within a specific class does not unduly penalize the degree of the design. However, the choice of weights is not immediate, as it is in the standard H_∞ paradigm [14]. The choice of weights that lead to acceptable controllers is, in itself, a non-convex optimization problem. Thus, one of the contributions of this paper is a relaxation of this non-convex problem into one which is quasi-convex, and thus solvable by standard methods. The methodology builds on a more fundamental question which forms a main theme of the paper, namely the characterization of all possible minimizers of weighted entropy functionals. The inverse problem of constructing weights for permissible minimizers is the basis for our new design theory. In a more general context, the results of this paper provide a solution to the longstanding open problem of determining which spectral zeros correspond to a certain desired shape of the interpolant.

This paper provides a considerable extension of the results presented in the conference paper [25]. The modified problem obtained by removing the *a priori* bound γ on the interpolants has been studied in [26, 28]. In fact, by allowing the upper bound γ to tend to infinity, the entropy optimization problem becomes an H_2 optimization problem, and the interpolants are then parameterized in terms of poles rather than in terms of spectral zeros.

A.10 Appendix

In the appendix we prove Theorem A.3 and some technical results needed in the paper.

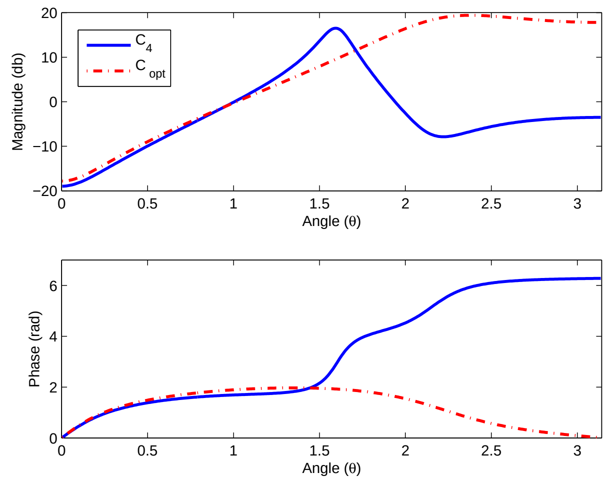

Figure A.7: Bode plots of controllers C_4, and C_{opt}

A.10.1 Controller complexity: Proof of Proposition A.2

Given the controller parameterization (A.8), we have $C = U/V$, where $U = U_0 + M_0 q$ and $V = V_0 + N_0 q$. Since $U, V \in H_\infty$, the number of the distinct poles of U and V (counted with multiplicity, including poles at zero and at ∞) is at least as large as the degree of C. From (A.11) and (A.12) we have

$$\begin{pmatrix} U \\ V \end{pmatrix} = \begin{pmatrix} \phi^* M_0 \\ \phi^* N_0 \end{pmatrix} F + \begin{pmatrix} -N_0^* \\ M_0^* \end{pmatrix},$$

and, since $M_0, N_0, \phi M_0^*, \phi N_0^* \in H_\infty$, any stable pole of U or V is a stable pole of F. However, U and V has no unstable poles, and therefore, the degree of C is bounded by the degree of F, as claimed.

A.10.2 Proof of Theorem A.3

The main ideas of the proof of Theorem A.3 are similar to those of Theorem 1 in [9], but some lemmas need to be modified to handle the present situation. For ease of reference, we retain (modulo a sign change) the notations of [9].

We start by showing that for any log-integrable weight Ψ, there exists a strictly contractive interpolant w with finite generalized entropy.

Lemma A.12. *Suppose that* $\Psi \geq 0$ *satisfies* $\log \Psi \in L_1(\mathbb{T})$*, and let* P *be the Pick matrix* (A.2)*. Then, if* $P > 0$*, there exists a* $w \in H_\infty$ *that satisfies* (A.1) *and* $\|w\|_\infty < 1$ *and such that* $\mathbb{K}_\Psi(w) < \infty$*.*

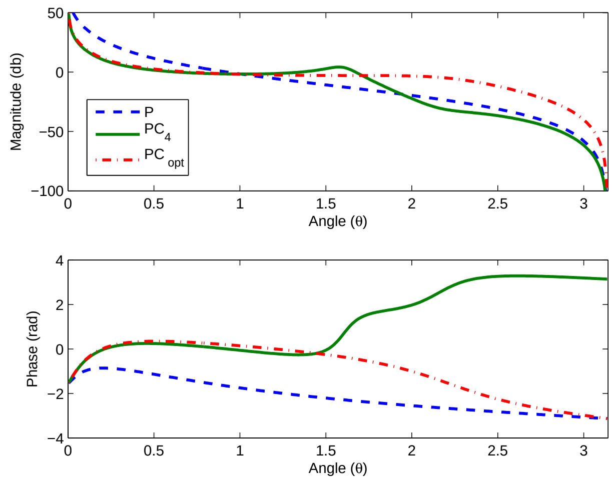

Figure A.8: Bode plots of P and of loop gains PC_4, and PC_{opt}

Proof. Let g_ϵ be the outer function such that $|g_\epsilon| = \sqrt{\frac{1}{1+\epsilon+\epsilon\Psi}}$, for $\epsilon > 0$. Since $g_\epsilon \to 1$ pointwise in $\mathbb{D}$ as $\epsilon \to 0$, and since $\text{Pick}(1) = P > 0$, it is possible to choose ϵ small enough so that $\text{Pick}(g_\epsilon) \geq 0$. Then it is possible to find a w which satisfies (A.1) and for which $|w| \leq |g_\epsilon|$ on $\mathbb{T}$. The calculation

$$\begin{aligned} \mathbb{K}_\Psi(w) &\leq \mathbb{K}_\Psi(g_\epsilon) \\ &= \int_\mathbb{T} \Psi \log\left(\frac{1+\epsilon+\epsilon\Psi}{\epsilon+\epsilon\Psi}\right) dm \\ &\leq \int_\mathbb{T} \Psi \left(\frac{1+\epsilon+\epsilon\Psi}{\epsilon+\epsilon\Psi} - 1\right) dm \\ &\leq \frac{2\pi}{\epsilon} \end{aligned}$$

shows that $\mathbb{K}_\Psi(w) < \infty$. □

With w fixed and satisfying the properties in Lemma A.12, we define

$$\begin{aligned} X &= \{v \mid \|w+\phi v\| \leq 1\}, \\ X^0 &= \{v \mid \|w+\phi v\| \leq 1, \mathbb{K}_\Psi(w+\phi v) < \infty\}, \end{aligned}$$

where ϕ is the Blaschke product (A.5).

Consider the convex functional $F : X \to [0,\infty]$ given by

$$F(v) = -\int_\mathbb{T} |\sigma|^2 \log(1-|w+\phi v|^2) dm \qquad \text{(A.29)}$$

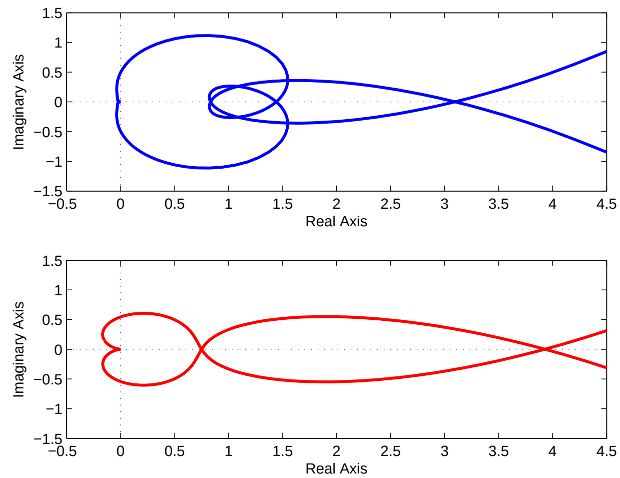

Figure A.9: Nyquist plots of the loop gains PC_4 (above) and PC_{opt} (below), respectively

where as previous, σ is outer such that $|\sigma|^2 = \Psi$

Lemma A.13. *There exists a set* $\mathfrak{L}$ *of continuous affine functionals on* H^∞ *of the form*

$$\lambda(v) = \lambda_0 + \int_{\mathbb{T}} \Re e(hv)dm,$$

with $\lambda_0 \in \mathbb{R}, h \in L^1$*, such that*

$$F(v) = \sup_{\lambda \in \mathfrak{L}} \lambda(v) \text{ for every } v \in X. \tag{A.30}$$

Proof. For all $a, x > 0$, $\log(x) \leq \log(a) + \frac{x-a}{a}$. Hence for all $|s| \leq 1, |z| \leq 1$, and $\epsilon \in (0, 1)$ we have

$$\begin{aligned} \log(1-(1-\epsilon)|s|^2) &\leq \log(1-(1-\epsilon)|z|^2) + (1-\epsilon)\frac{|z|^2 - |s|^2}{1-(1-\epsilon)|z|^2} \\ &\leq \log(1-(1-\epsilon)|z|^2) + 2(1-\epsilon)\frac{|z|^2 - \Re e(\bar{z}s)}{1-(1-\epsilon)|z|^2} \end{aligned}$$

with equality if $s = z$. Let $\Psi_\epsilon := \min(\Psi, 1/\epsilon)$. By monotone convergence, $\mathbb{K}_{\Psi_\epsilon}((1-$

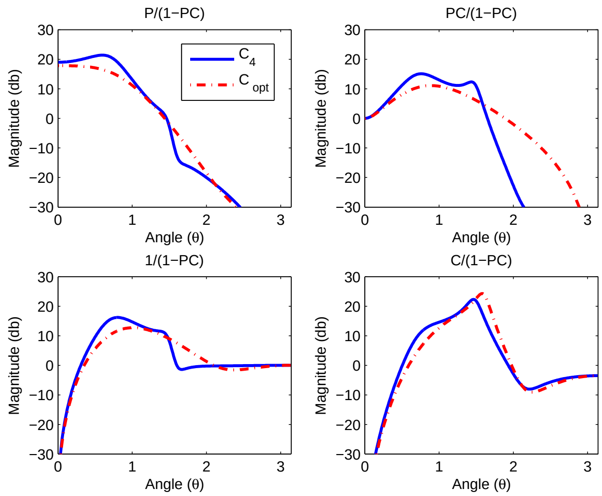

Figure A.10: The four closed-loop transfer functions of $\Pi_{P//C_4}$ and $\Pi_{P//C_{\text{opt}}}$

$\epsilon)f) \to \mathbb{K}_\Psi(f)$ as $\epsilon \to 0$. Consequently,

$$
\begin{aligned}
F(v) &= \sup_{\epsilon\in(0,1)} -\int_{\mathbb{T}} \Psi_\epsilon \log(1-(1-\epsilon)|w+\phi v|^2)dm \\
&\geq \sup_{u\in X}\sup_{\epsilon\in(0,1)} -\int_{\mathbb{T}} \Psi_\epsilon \Bigg(\log(1-(1-\epsilon)|w+\phi u|^2) \\
&\quad +2(1-\epsilon)\frac{|w+\phi u|^2 - \Re e(\overline{(w+\phi u)}(w+\phi v)}{1-(1-\epsilon)|w+\phi u|^2}\Bigg)dm,
\end{aligned}
$$

where the outer infimum is achieved at $u = v$. Therefore, by defining $\mathcal{L}$ to be the class of affine functionals on H_∞ defined via

$$
\lambda_0 = -\int_{\mathbb{T}} \Psi_\epsilon \Bigg(\log(1-(1-\epsilon)|w+\phi u|^2) + 2(1-\epsilon)\frac{|w+\phi u|^2 - \Re e((\overline{w+\phi u})w)}{1-(1-\epsilon)|w+\phi u|^2}\Bigg)dm
$$

and

$$
h = \Psi_\epsilon \frac{2(1-\epsilon)(\overline{w+\phi u})\phi}{1-(1-\epsilon)|w+\phi u|^2},
$$

for all $\epsilon \in (0,1)$ and $u \in X$, the representation (A.30) follows. □

This leads to the main lemma, the proof of which follows verbatim from [9, p. 974] using Lemma A.13.

Main Lemma A.14. *The optimization problem* (A.4) *has a unique minimizer on* X.

Lemma A.15. *Let* $f = w + \phi v$*, where* $v = \arg\min(F)$*. Then*

$$\frac{|\sigma f|^2}{1 - |f|^2} \in L_1(\mathbb{T}).$$

Proof. Consider for $t \in (0, 1)$, the differentiable function $\psi : t \mapsto F((1-t)v)$ with derivative

$$\dot{\psi}(t) = \int_{\mathbb{T}} \varphi_t dm,$$

where

$$\varphi_t := |\sigma|^2 \frac{\Re e\left\{(\bar{f} + t(\bar{w} - \bar{f}))(f - w)\right\}}{1 - |f + t(w - f)|^2}.$$

Since F has a minimum at v, $\dot{\psi}(t) \geq 0$ for small t. Choose a γ such that $\frac{\|w\|_\infty + 1}{2} \leq \gamma < 1$, and let I_1 and I_2 be the subintervals of $\mathbb{T}$ where $|f| < \gamma$ and $|f| \geq \gamma$, respectively. On I_1, the denominator of φ_t is bounded, and, since $\int_{\mathbb{T}} |\sigma f|^2 dm \leq \mathbb{K}_\Psi(f) < \infty$ and $\int_{\mathbb{T}} |\sigma w|^2 dm \leq \mathbb{K}_\Psi(w) < \infty$, we have that

$$\left| \int_{I_1} \varphi_t dm \right|$$

is uniformly bounded for $t \in (0, 1)$, while, on I_2, $\varphi_t < 0$ for sufficiently small t. Therefore, to say that $\frac{|\sigma f|^2}{1-|f|^2} \notin L_1(\mathbb{T})$ is to say that

$$\int_{I_2} \varphi_t dm \to -\infty,$$

contradicting the nonnegativity of $\dot{\psi}(t)$ for small t. □

Since σ is log-intergrable and $f \in H_\infty$ with $\|f\| \leq 1$, the fact that $\frac{|\sigma f|^2}{1-|f|^2} \in L_1(\mathbb{T})$ implies that $\log(1 - |f|^2) \in L_1$. Therefore $\log \frac{|\sigma|^2}{1-|f|^2} \in L_1(\mathbb{T})$ and there is a unique outer factor a with $a(0) > 0$ such that

$$|a|^2 = \frac{|\sigma|^2}{1 - |f|^2},$$

and, if we define $b := fa$ (in H_2 due to Lemma A.15), a and b satisfies

$$\begin{aligned} f &= \frac{b}{a}, \\ |\sigma|^2 &= |a|^2 - |b|^2. \end{aligned}$$

Next, we show that $b \in \mathcal{K}$.

Lemma A.16. *Let $f = w + \phi v$, where $v = \arg\min(F)$, and let f_0 be the outer part of f. Then*

$$\frac{|\sigma|^2 f_0 f^* \phi}{1-|f|^2} \in zH_1(\mathbb{D}).$$

Proof. From the Lemma A.15 we have that $\frac{|\sigma f|^2}{1-|f|^2} \in L_1(\mathbb{T})$, and, since $\log|\sigma| \in L^1$, it follows that $\log(1-|f|^2) \in L^1$. Hence there is $g \in H_2$ such that $1-|f|^2 = |g|^2$. For any $h \in H^\infty$ with $\|h\|_\infty \leq 1$, we have that $v + tg^2 f_0 h \in X_0$ for $t \in (-\epsilon, \epsilon)$ and ϵ sufficiently small. In fact, since $\Re e\{\bar{f}\phi t g^2 f_0 h\} \leq |t||f|^2$, we have

$$1 - |f + t\phi g^2 f_0 h|^2 \geq |g|^2(1 - 2|t||f|^2 - t^2|f|^4),$$

and hence

$$\begin{aligned} F(v + tg^2 f_0 h) &= -\int_{\mathbb{T}} |\sigma|^2 \log(1 - |f + t\phi g^2 f_0 h|^2) dm \\ &\leq F(v) - \int_{\mathbb{T}} |\sigma|^2 \log(1 - 2|t||f|^2 - t^2|f|^4) dm. \end{aligned}$$

In fact, since $-\log(1-x) \leq 2x$ for $x \in (0, \frac{1}{2})$, we have for $t \in (-\frac{1}{6}, \frac{1}{6})$ that

$$\begin{aligned} -\int_{\mathbb{T}} |\sigma|^2 \log(1 - 2|t||f|^2 - t^2|f|^4) dm &\leq -\int_{\mathbb{T}} |\sigma|^2 \log(1 - 3|t||f|^2) dm \\ &\leq 6|t| \int_{\mathbb{T}} |\sigma|^2 |f|^2 dm < \infty. \end{aligned}$$

This proves that $v + tg^2 f_0 h \in X_0$ for $t \in (-\frac{1}{6}, \frac{1}{6})$. Since $v = \arg\min(F)$, the derivative of F must be zero at v in the directions $g^2 f_0 h$ for arbitrary $h \in H_\infty$ with $\|h\|_\infty \leq 1$; i.e.,

$$\int_{\mathbb{T}} |\sigma|^2 \frac{\Re e\{f^* \phi g^2 f_0 h\}}{1-|f|^2} dm = 0$$

for all $h \in H_\infty$ in the unit ball. Therefore,

$$|\sigma|^2 \frac{f^* \phi g^2 f_0}{1-|f|^2} \in zH_1(\mathbb{D}),$$

so, since g is outer and $|\sigma|^2 \frac{f^* \phi f_0}{1-|f|^2} \in L_1$, Lemma A.16 follows. ☐

By Lemma A.15, we have that $b = af = \frac{\sigma f}{g} \in H_2$. By Lemma A.16, $\frac{b^* \sigma f_0 \phi}{g} = zq$ where $q \in H_1$, and hence $\frac{b^* \phi}{z} = \frac{qg}{\sigma} \in H_2$. Since $b \in H_2$ and $\frac{b^* \phi}{z} \in H_2$ it follows that $b \in \mathcal{K}$. Finally, since $a = \frac{\sigma}{g}$, the minimizing interpolant f is of the form stated in Theorem A.3.

We have thus established that the minimizer exists and satisfies (i), (ii), and (iii). It remains to prove that there is only one function $f \in \mathcal{S}$ satisfying (i), (ii), and (iii). This follows directly from the arguments on page 977 in [9] and by noting that

$$|\sigma|^2 \frac{|f_1^*(f_1 - f_2)|}{1 - |f_1|^2} \leq \frac{|\sigma f_1|^2}{1 - |f_1|^2} + \frac{|\sigma f_2|^2}{1 - |f_2|^2} \in L_1$$

and hence that

$$|\sigma|^2 \frac{2\Re e\{f_1^*(f_1 - f_2)\}}{1 - |f_1|^2} \in L_1.$$

This concludes the proof of Theorem A.3.

A.10.3 A lemma on continuity

Lemma A.17. *Let* $\Psi \in \mathcal{M}$ *and* $f = \frac{b}{a} = \varphi(\Psi)$*, and let* $\Psi_k \in \mathcal{M}$ *and* $f_k = \frac{b_k}{a_k} = \varphi(\Psi_k)$ *for* $k = 1, 2, \ldots$*. Then, if*

$$\| \log(\Psi) - \log(\Psi_k)\| \to 0 \quad \textit{as } k \to \infty,$$

$b_k \to b$ *as* $k \to \infty$*.*

Proof. From the proof of the main theorem, we see that

$$b = \frac{f\sigma}{g},$$

with σ the outer factor of Ψ and g the outer factor of $1 - |f|^2$. Similarly $b_k = \frac{f_k \sigma_k}{g_k}$ where σ_k, g_k are outer, $|\sigma_k|^2 = \Psi_k$ and $|g_k|^2 = 1 - |f_k|^2$.

As $\| \log(\Psi) - \log(\Psi_k)\| \to 0$, $f_k \to f$ in H_2 (Proposition A.10). From this it follows that $1 - |f_k|^2 \to 1 - |f|^2$ in L_1 and hence that $g_k \to g$ in H_2 (see [5] and note that $\{1 - |f_k|^2\}_k$ are log-integrable). Clearly $\sigma_k \to \sigma$ in H_2.

Since $b, b_k \in \mathcal{K}$, $b = \frac{\beta}{\tau}$, $b_k = \frac{\beta_k}{\tau}$ with $\beta, \beta_k \in \mathrm{Pol}(n)$, we therefore have

$$g_k \beta_k = f_k \sigma_k \tau.$$

The convergence $b_k \to b$ now follows, since $g_k \to g$ $f_k \to f$ and $\sigma_k \to \sigma$ coefficientwise (weakly) and g is not identically zero. □

A.11 Bibliography

[1] J. Agler and J.E. McCarthy. *Pick Interpolation and Hilbert Function Spaces*. American Mathematical Society, 2002.

[2] B.D.O. Anderson. Continuity of the spectral factorization operation. *Mat. Apl. Comput.*, 4(2):139–156, 1985.

[3] P. Apkarian and D. Noll. Nonsmooth optimization for multiband frequency domain control design. *Automatica*, 43:724–731, 2007.

[4] J. Ball, I. Gohberg, and L. Rodman. *Interpolation of rational matrix functions*, volume 45 of *Operator Theory: Advances and Applications*. Boston, Birkhaüser, 1990.

[5] S. Barclay. Continuity of the spectral factorization mapping. *Journal of the London Mathematical Society*, 70:763–779, 2004.

[6] S. A. Buddie, T. T. Georgiou, Ü. Özgüner, and M. C. Smith. Flexible structure experiments at JPL and WPAFB: H_∞ controller design. *International Journal of Control*, 58:1–19, 1993.

[7] C. I. Byrnes, P. Enqvist, and A. Lindquist. Cepstral coefficients, covariance lags and pole-zero models for finite data strings. *IEEE Trans. SP*, 677-693, 2001.

[8] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[9] C. I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski. Generalized interpolation in H-infinity with a complexity constraint. *Transactions of the American Mathematical Society*, 358:965–987, 2006.

[10] C. I. Byrnes, S. V. Gusev, and A. Lindquist. A convex optimization approach to the rational covariance extension problem. *SIAM J. Contr. and Optimiz.*, 37:211–229, 1998.

[11] C. I. Byrnes, S. V. Gusev, and A. Lindquist. From finite covariance windows to modeling filters: A convex optimization approach. *SIAM Review*, 43, 2001.

[12] C. I. Byrnes, A. Lindquist, S. V. Gusev, and A. S. Matveev. A complete parameterization of all positive rational extensions of a covariance sequence. *IEEE Trans. Automat. Control*, 40:1841–1857, 1995.

[13] H. Dym and I. Gohberg. On maximum entropy interpolants and maximum determinant completions of associated Pick matrices. *Integral Equations Operator Theory*, 23(1), March 1995.

[14] B.A. Francis. *A Course in H_∞ Control Theory*. Springer Verlag, New York, 1987.

[15] P. Gahinet and P. Apkarian. A linear matrix inequality approach to H^∞ control. *Intern. J. of Robust and Nonlinear control*, 4:421–448, 1994.

[16] J.B. Garnett. *Bounded analytic Functions*. Academic Press, 1981.

[17] T. T. Georgiou. Realization of power spectra from partial covariance sequences. *IEEE Trans. Acoustics, Speech and Signal Processing*, 35:438–449, 1987.

[18] T. T. Georgiou. A topological approach to Nevanlinna-Pick interpolation. *SIAM J. on Mathematical Analysis*, 18(5):1248–1260, September 1987.

[19] T.T. Georgiou. *Partial Realization of Covariance Sequences*. PhD thesis, CMST, University of Florida, Gainesville, 1983.

[20] T.T. Georgiou and A. Lindquist. Kullback-Leibler approximation of spectral density functions. *IEEE Trans. on Information Theory*, 49(11), November 2003.

[21] T.T. Georgiou and A. Lindquist. Remarks on control design with degree constraint. *IEEE Transactions on Automatic Control*, 51:1150–1156, July 2006.

[22] T.T. Georgiou and M. Smith. Optimal robustness in the gap metric. *IEEE Trans. on Automatic Control*, 35(6), June 1990.

[23] T. Iwasaki and S. Hara. Generalized kyp lemma: unified frequency domain inequalities with design applications. *IEEE Trans. on Automatic Control*, 50(1):41–59, January 2005.

[24] R.E. Kalman. When is a linear control system optimal? *Journal of Basic Engineering*, pages 51–60, March 1964.

[25] J. Karlsson, T. Georgiou, and A Lindquist. The inverse problem of analytic interpolation with degree constraint. In *IEEE Proc. CDC*, San Diego, 2006.

[26] J. Karlsson and A. Lindquist. Stable rational approximation in the context of interpolation and convex optimization. In *46th Conference on Decision and Control*, pages 4353–4360, New Orleans, 2007.

[27] J. Karlsson and A Lindquist. On degree-constrained analytic interpolation with interpolation points close to the boundary. *Provisionally accepted to IEEE Transactions of Automatic Control*, 2008.

[28] J. Karlsson and A. Lindquist. Stability-preserving rational approximation subject to interpolation constraints. *IEEE Transactions on Automatic Control*, 2008.

[29] D. Macfarlane and K. Glover. A loop-shaping design procedure using H_∞ synthesis. *IEEE Trans. Automat. Control*, 37(6):759–769, June 1992.

[30] D. Mustafa and K. Glover. *Minimum Entropy H-infinity Control*. Springer-Verlag, Berlin Heidelberg, 1990.

[31] R. Nagamune. Closed-loop shaping based on the Nevanlinna-Pick interpolation with a degree bound. *IEEE Trans. Automat. Control*, 49:300–305, 2004.

[32] R. Nagamune and A. Blomqvist. Sensitivity shaping with degree constraint by nonlinear least-square optimization. *Automatica*, 41(7):1219–1227, 2005.

[33] R.E. Skelton, T. Iwasaki, and K.M. Grigoriadis. *A Unified Algebraic Approach to Linear Control Design*. Taylor Francis, London, 1997.

[34] M. Vidyasagar. *Control Systems Synthesis: A Factorization Approach*. MIT Press, Cambridge, 1995.

[35] M. Vidyasagar and H. Kimura. Robust controllers for uncertain linear multivariable systems. *Automatica*, 22(1):85–94, January 1986.

[36] G. Vinnicombe. Frequency domain uncertainty and the graph topology. *IEEE Trans. on Automat. Control*, 38:1371–1383, 1993.

[37] G. Zames. Feedback and optimal sensitivity: Model reference transformations, multiplicative seminorms and approximate inverses. *IEEE Trans. Automat. Contr.*, 26:301–320, 1981.

[38] K. Zhou. Frequency-weighted l_∞ norm and optimal Hankel norm model reduction. *IEEE Trans. Automat. Control*, 40:1687–1699, 1995.

On Degree-Constrained Analytic Interpolation with Interpolation Points Close to the Boundary

Johan Karlsson and Anders Lindquist

Abstract

In the recent article [4] a theory for complexity-constrained interpolation of contractive functions is developed. In particular, it is shown that any such interpolant may be obtained as the unique minimizer of a (convex) weighted entropy gain. In this paper we study this optimization problem in detail and describe how the minimizer depends on weight selection and on interpolation conditions. We first show that, if, for a sequence of interpolants, the values of the entropy gain of the interpolants converge to the optimum, then the interpolants converge in H_2, but not necessarily in H_∞. This result is then used to describe the asymptotic behavior of the interpolant as an interpolation point approaches the boundary of the domain of analyticity. For loop shaping to specifications in control design, it might at first seem natural to place strategically additional interpolation points close to the boundary. However, our results indicate that such a strategy will have little effect on the shape. Another consequence of our results relates to model reduction based on minimum-entropy principles, where one should avoid placing interpolation points too close to the boundary.

B.1 Introduction

Many important engineering problems lead to analytic interpolation, where the interpolant represents a transfer function of, for example, a feedback control system or a filter and therefore is required to be a rational function of bounded degree. In recent years, a complete theory of analytic interpolation with degree constraint has been developed, which provides complete smooth parameterizations of whole classes of such interpolants in terms of a weighting function belonging to a finite-dimensional space, as well as convex optimization problems for determining them; see [3, 4] and references therein.

This theory provides a framework for tuning an engineering design based on analytic interpolation to satisfy additional design specification without increasing the degree of the transfer function. Occasionally the number of tuning parameters is too small to satisfy the design specifications, and then the parameter space needs to be enlarged by increasing the degree bound. In [11] this was done by adding new interpolation conditions, often close to the boundary.

In this paper we present some negative results concerning this strategy and explain why, after all, the solution in [11] is satisfactory. We show that unless the weighting function is changed, adding new interpolation points close to the boundary will have little effect on the interpolant. We illustrate this by analyzing a simple example from robust control.

We also show that interpolation conditions close to the unit disc have little effect on the minimum-entropy solution and can thus be discarded (Remark B.2). Recently some procedures for model reduction based on the minimum-entropy solution have been proposed [2, 12], which amount to interpolating in the mirror images of selected spectral zeros. Our results suggest (at least for bounded-real interpolants) that dominant spectral zeros close to the boundary should not be selected in this procedure. However, by choosing more general weights [8], this situation can be avoided.

In Section B.2 we begin by reviewing some pertinent results from [4] amplified with a generalization from the more recent paper [9]. Then, in Section B.3, we provide a motivation example from robust control, which is then revisited in Section B.5 after having presented the main results in Section B.4. Some proofs are deferred to Section B.6.

B.2 Background

Consider the classical Nevanlinna-Pick problem of finding a function f in the Schur class

$$\mathcal{S} := \{f \in H_\infty(\mathbb{D}) \,:\, \|f\|_\infty \leq 1\}$$

that satisfies the interpolation condition

$$f(z_k) = w_k, \quad k = 0, 1, \ldots, n, \tag{B.1}$$

where $(z_k, w_k), k = 0, 1, \ldots, n$, are given pairs of points in the open unit disc $\mathbb{D} := \{z : |z| < 1\}$. It is well-known that such an $f \in \mathcal{S}$ exists if and only if the Pick matrix

$$P := \left[\frac{1 - w_k \bar{w}_\ell}{1 - z_k \bar{z}_\ell}\right]_{k,\ell=0}^{n} \tag{B.2}$$

is positive semi-definite, and that the function f is unique if and only if the matrix P is singular. In the latter case f is a Blaschke product of degree equal to the rank of P. Here we shall take P to be positive definite, in which case there are infinitely many solutions to the Pick problem. A complete parameterization of the solutions of this so called Nevanlinna-Pick interpolation problem was given by Nevanlinna (see, e.g. [1]) in 1929. The parameterization is in terms of a linear fractional transformation centered around a rational solution of degree n, known as the *central solution*.

In a research program leading to [3, 4], the subset of all solutions of the Nevanlinna-Pick problem that are rational of degree at most n were parameterized. Most engineering problems require such degree constraints, which completely alter the basic mathematical problem. More precisely, let $\mathcal{K}$ be the space of all functions

$$\mathcal{K} = \left\{ \frac{p(z)}{\tau(z)} \,:\, p(z) \in \text{Pol}(n),\ \tau(z) = \prod_{k=0}^{n} (1 - \bar{z}_k z) \right\},$$

where Pol(n) denotes the set of polynomial of degree at most n. Clearly $\mathcal{K}$ is a subspace of the Hardy space $H_2(\mathbb{D})$. Moreover, let $\mathcal{K}_0$ be the the subset of all $f \in \mathcal{K}$ such that $p(z)$ has all its roots in the complement of $\mathbb{D}$ and $p(0) > 0$. In this (rational) context, Theorem 1 in [4] can be stated in the following way.

Theorem B.1. *Suppose that the Pick matrix* (B.2) *is positive definite. Let σ be an arbitrary function in $\mathcal{K}_0$. Then there exists a unique pair $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ such that*

(i) $f = b/a \in \mathcal{S}$

(ii) $f(z_k) = w_k,\ k = 0, 1, \ldots, n$*, and*

(iii) $|a|^2 - |b|^2 = |\sigma|^2$ *a.e. on* $\mathbb{T} := \{z \,:\, |z| = 1\}$.

Conversely, any pair $(a, b) \in \mathcal{K}_0 \times \mathcal{K}$ satisfying (i) and (ii) determines, via (iii), a unique $\sigma \in \mathcal{K}_0$.

Consequently, the solutions (a, b) corresponding to interpolants of degree at most n are completely parameterized by the zeros of $\sigma \in \mathcal{K}_0$; i.e., the n-tuples $\{\lambda_1, \ldots, \lambda_n\}$ of complex number in the complement of $\mathbb{D}$; these are called the *spectral zeros*. For each such choice of spectral zeros, the corresponding interpolant $f \in \mathcal{S}$ can be determined by minimizing the strictly convex functional $\mathbb{K}_\Psi : \mathcal{S} \to \mathbb{R} \cup \{\infty\}$, given by

$$\mathbb{K}_\Psi(f) = -\int_{\mathbb{T}} \Psi \log(1 - |f|^2) dm(z),$$

over the class of interpolants, where $\Psi := |\sigma|^2$ and m is the normalized Lebesgue measure on $\mathbb{T}$. In fact, in the present context, Theorem 5 in [4] can can be stated as follows.

Theorem B.2. *Suppose that the Pick matrix* (B.2) *is positive definite. Let σ be an arbitrary function in $\mathcal{K}_0$, and set $\Psi := |\sigma|^2$. Then the functional $\mathbb{K}_\Psi$ has a unique minimizer in the class of functions that satisfy the interpolation conditions* (B.1)*, and this minimizer is precisely the unique function $f \in \mathcal{S}$ satisfying conditions (i), (ii) and (iii) in Theorem B.1.*

Remark B.1. When $z_0 = 0$, then the central solution corresponds to $\Psi \equiv 1$. The corresponding functional $\mathbb{K}_1$ is the usual entropy gain, and the central solution is therefore equal to the *minimum entropy solution* (see, e.g. [10]). Then $\sigma \equiv 1 \in \mathcal{K}_0$, and hence the generic degree of the minimum entropy solution is n, and the corresponding spectral zeros are located at the conjugate inverses (mirror images in unit circle) of $\{z_k\}_{k=1}^n$; all in harmony with Theorems B.1 and B.2. If zero is not an interpolation point, then $1 \notin \mathcal{K}_0$ and the generic degree of the minimum entropy solution is instead $n + 1$.

By varying Ψ we can tune the interpolant without increasing its degree. In the context of our motivating example in the next section, this amounts shaping the sensitivity function without increasing the McMillan degree of the closed-loop system. By adding interpolation conditions we can increase the number of tuning parameters at the cost of increased degree of the interpolant.

Theorems B.1 and B.2 can be generalized to the case that Ψ is an arbitrary log-integrable function on $\mathbb{T}$. This was done in the following way in [9].

Theorem B.3. *Suppose that the Pick matrix* (B.2) *is positive definite and that* $\Psi = |\sigma|^2$ *is a log-integrable nonnegative function on the unit circle, where* σ *is analytic but need not belong to* $\mathcal{K}_0$. *Then* f *is the minimizer of* $\mathbb{K}_\Psi$ *in the class of functions that satisfy the interpolation conditions* (B.1) *if and only if the following three conditions hold:*

(i) $f = b/a \in \mathcal{S}$ *where* $b \in \mathcal{K}$ *and* a *is outer,*

(ii) $f(z_k) = w_k$ *for* $k = 0, \dots, n$,

(iii) $|a|^2 - |b|^2 = |\sigma|^2$.

Any such minimizer is necessarily unique.

This allows for shaping the interpolant without the constraint that σ belong to $\mathcal{K}_0$, but at the expense of increased degree; for a precise statement, see [9]. An interesting special case of this is when $\Psi \in C(\mathbb{T})_+$, i.e. Ψ is positive and continuous on the unit circle.

The theory described above allows us to choose an interpolant that satisfies additional design specifications. In fact, the map from σ to (a, b) defined by Theorem B.1 is smooth [5, 6], and hence a given design can be tuned via Ψ to smoothly change the interpolant. An obvious first choice of Ψ is to make it large in frequency bands where $|f|$ needs to be small. This paper is an attempt to gain understanding of the underlying function theory involved in tuning the interpolant. In the paper [9] we have derived a *systematic* procedure for shaping interpolants based on design specifications.

B.3 A motivating example

The purpose of this paper is to show how the interpolant changes as the weight Ψ is changed and as additional interpolation points are introduced, especially close to the boundary of $\mathbb{D}$. To illustrate this point, we consider an example on sensitivity shaping in robust control from [11]. Figure B.1 depicts a feedback system with u denoting the control input to the plant

$$G(z) = \frac{1}{z - 1.05}$$

to be controlled, d represents a disturbance, and y is the resulting output, which is fed back through a compensator $K(z)$ to be designed. The goal is to determine a controller $K(z)$

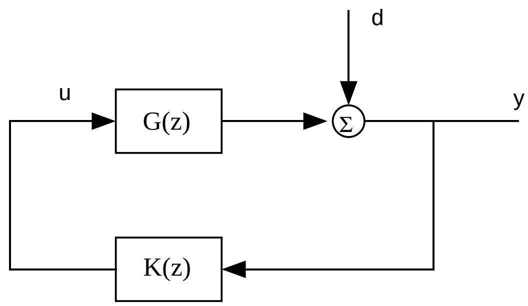

Figure B.1: A feedback system.

so that the feedback system in Figure B.1 satisfies the design specifications

$$|S(e^{i\theta})| < 2.0 \sim \quad 6.02dB \qquad 0 \ \leq \theta \leq \pi \tag{B.3}$$
$$|S(e^{i\theta})| < 0.1 \sim -20.1dB \qquad 0 \ \leq \theta \leq 0.3 \tag{B.4}$$
$$|T(e^{i\theta})| < 0.5 \sim -6.02dB \quad 2.5 \leq \theta \leq \pi \tag{B.5}$$

in terms of the sensitivity function $S = (1 - GK)^{-1}$ and the complementary sensitivity function $T = 1 - S$.

The plant $G(z)$ has one unstable pole at $z = 1.05$ and one non-minimum phase zero at $z = \infty$. It follows from H_∞ control theory (see, e.g. [7]) that the feedback system is internally stable if and only if the sensitivity function $S(z)$, the transfer function from d to y, is analytic in $\mathbb{D}^C := \{z \, : \, |z| > 1\}$, the complement of the closed unit disc, and satisfies the interpolation conditions

$$S(1.05) = 0, \quad S(\infty) = 1.$$

By the design specification (B.3), the interpolants are required to satisfy $\|S\|_\infty \leq \gamma := 2$. Setting

$$f(z) = \frac{1}{2} S(z^{-1}), \tag{B.6}$$

f fits into the framework of Theorem B.1 with interpolation conditions

$$f(0.9524) = 0, \quad f(0) = \frac{1}{2}. \tag{B.7}$$

Since $n = 1$, there exists a one-parameter family of degree-one interpolants satisfying $\|f\|_\infty \leq 1$ that may be parametrized by its corresponding spectral zero λ. Figure B.2 shows the solutions $S(z) = 2f(z^{-1})$ as $1/\lambda$ varies from -1 to 1 with the grid 0.2.

Clearly none of these designs satisfies the specifications. Therefore, following Nagamune [11], we add the interpolation conditions

$$f(-0.9901) = \frac{1}{2} \tag{B.8}$$

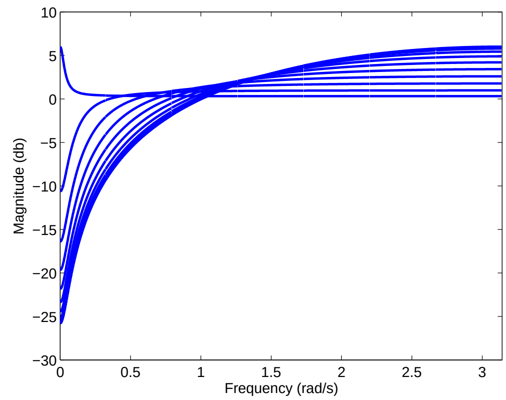

Figure B.2: *The degree-one sensitivity functions corresponding to spectral-zero selections with* $1/\lambda$ *between* -1 *to* 1 *with grid* 0.2*.*

and

$$f(0.9901e^{\pm 0.3i}) = 0. \tag{B.9}$$

Here (B.8), motivated by the design specification (B.5), ensures, via (B.6), that $T(1.01) = 0$, while (B.9), motivated by the design specification (B.4), ensures that $S(1.01e^{\pm 0.3i}) = 0$. The number of interpolation conditions adds up to $n + 1 = 5$, and therefore Theorem B.1 allows for parameterizing all solutions of degree $n = 4$ by choosing n spectral zeros. As in [11], we choose the spectral zeros in $0.97e^{\pm 0.55i}$ and $0.9e^{\pm 1.55i}$, which corresponds to the weight

$$\Psi_{\mathrm{N}} = \left| \frac{(1 - 0.9e^{1.55i}z)(1 - 0.9e^{-1.55i}z)(1 - 0.97e^{0.55i}z)(1 - 0.97e^{-0.55i}z)}{(z - 1.05)(z + 1.01)(z - 1.01e^{0.3i})(z - 1.01e^{-0.3i})} \right|^2 . \tag{B.10}$$

The corresponding sensitivity function S_{N}, depicted in Figure B.3 together with the weight Ψ_{N}, satisfies the design specifications (B.3)-(B.5). The interpolation points and spectral zeros for this design are depicted in Figure B.4.

From the plots in Figure B.3 one first notices that in the example where a large weight in the low frequency region is used, the magnitude of the sensitivity is low. This seems to be intuitive since the high weight in the entropy functional penalizes the sensitivity more in that region than in others. However, the weight is also large in the high frequency area, and in this case there is no significant change in the sensitivity.

In order to understand the effects of interpolation points and spectral zeros in this design, in the next section we develop results for interpolation points close to the unit circle. Then, in Section B.5 we revisit the example above.

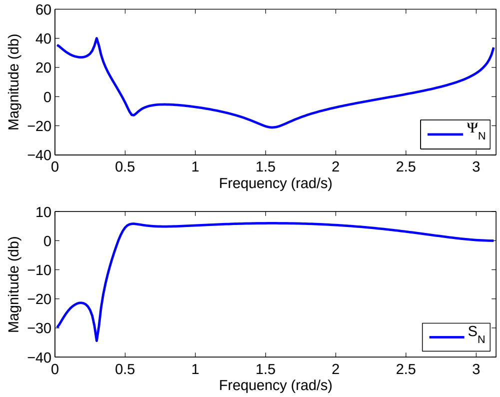

Figure B.3: *The weight Ψ_N (above) and the magnitude of the sensitivity function S_N (below).*

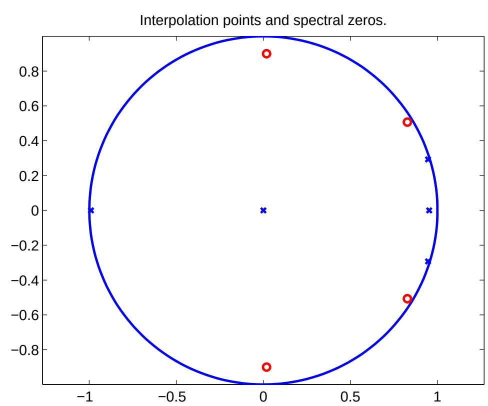

Figure B.4: *The intepolation points ($\times$) and the mirror images ($\circ$) of the spectral zeros corresponding to S_N.*

B.4 Main results

As the example of Section B.3 suggests, we need to investigate how the interpolant changes as additional interpolation points are introduced close to the unit circle. The following theorem is one of our main results.

Theorem B.4. *Let $\Psi = |\sigma|^2 \in C(\mathbb{T})_+$, where $\sigma \in \mathcal{K}_0$, and let $\hat{f}$ be the minimizer of $\mathbb{K}_\Psi(f)$ subject to the interpolation conditions*

$$f(z_k) = w_k, \quad k = 0, 1, \ldots, n.$$

Moreover, given $|w| < 1$, let f_λ be the minimizer of $\mathbb{K}_\Psi(f)$ subject to

$$f(z_k) = w_k, \quad k = 0, 1, \ldots, n, \quad f(\lambda) = w.$$

Then $f_\lambda \to \hat{f}$ in H_2 as $|\lambda| \to 1$.

This theorem indicates that adding interpolation conditions close to the unit circle will not affect the design in any important way unless we also change the weighting function Ψ. For the proof we first need to show that if the generalized entropy of interpolants converge to the optimum, then the interpolants converge to the optimal interpolant in H_2. The following theorem is proved in Section B.6.

Theorem B.5. *Let $\Psi \in C(\mathbb{T})_+$, and let $\hat{f}$ be the minimizer of $\mathbb{K}_\Psi(f)$ subject to $f(z_k) = w_k$, $k = 0, 1, \ldots, n$. If f_ℓ satisfies $f_\ell(z_k) = w_k$, $k = 0, 1, \ldots, n$, and $\mathbb{K}_\Psi(f_\ell) \to \mathbb{K}_\Psi(\hat{f})$, then $f_\ell \to \hat{f}$ in H_2.*

It should be noted that this result could not be strengthened to H_∞ convergence. A counterexample could be constructed by noting that $\mathbb{K}_\Psi(f + \alpha\chi_{E_\ell}) \to \mathbb{K}_\Psi(f)$ if $m(E_\ell) \to 0$. Here χ denotes the characteristic function and α is a scalar such that $0 < |\alpha| < 1 - \|f\|_\infty$. But $\|\alpha\chi_{E_\ell}\|_\infty = \alpha$ for all ℓ. This argument works equally well for $f + g_\ell$, where $g_\ell \in \phi H_2$ and $|g_\ell|$ is an appropriate approximation of χ_{E_ℓ}.

A second step in proving Theorem B.4 is to investigate how the interpolant changes as the data is transformed under a Möbius transformation. For $\lambda \in \mathbb{D}$, let b_λ be the Blaschke factor

$$b_\lambda(z) = \frac{\lambda - z}{1 - \bar{\lambda} z}.$$

Then the following proposition tells us how the entropy is changed as the range is transformed by a Möbius transformation.

Proposition B.6. *The map $\rho(\cdot, \lambda, \Psi) : \mathcal{S} \to \mathbb{R}$ defined by*

$$\rho(f, \lambda, \Psi) = \int_{\mathbb{T}} \Psi \log \frac{|1 - \bar{\lambda} f|^2}{1 - \bar{\lambda}\lambda} dm(z)$$

is continuous, and

$$\mathbb{K}_\Psi(b_\lambda(f)) = \mathbb{K}_\Psi(f) + \rho(f, \lambda, \Psi).$$

Moreover, if $\Psi = |\sigma|^2$ where $\sigma \in \mathcal{K}_0$, then $\rho(f_1, \lambda, \Psi) = \rho(f_2, \lambda, \Psi)$, whenever $f_1(z_k) = f_2(z_k)$ for $k = 0, 1, \ldots, n$.

Proof. First part is trivial, second part follows from [4, p. 8 and Lemma 10]. □

As a corollary we have the following proposition, which tells us that the solution obtained from the transformed data is the solution transformed with the same transformation.

Proposition B.7. *Let $\sigma \in \mathcal{K}_0$, and let f be the minimizer of $\mathbb{K}_{|\sigma|^2}(f)$ subject to $f(z_k) = w_k, k = 0, 1, \ldots, n$, as prescribed by Theorem B.1. Then $g = b_\lambda(f)$ is the minimizing interpolant corresponding to the same σ of the analytic interpolation problem $g(z_k) = b_\lambda(w_k), k = 0, 1, \ldots, n$.*

A simple proof of Proposition B.7, derived directly from Theorem B.1 using basic principles, is given in Section B.6.

To conclude the proof of Theorem B.4, we first prove a version in which the interpolation value w equals zero.

Theorem B.8. *Let $\Psi \in C(\mathbb{T})_+$, and let $\hat{f}$ be the minimizer of $\mathbb{K}_\Psi(f)$ subject to the interpolation conditions $f(z_k) = w_k$, $k = 0, 1, \ldots, n$. Moreover, let f_λ be the minimizer of $\mathbb{K}_\Psi(f)$ subject to $f(z_k) = w_k$, $k = 0, 1, \ldots, n$ and $f(\lambda) = 0$. Then $f_\lambda \to \hat{f}$ in H_2 as $|\lambda| \to 1$.*

In Theorems B.4 and B.8, f_λ may not exist for certain $\lambda \in \mathbb{D}$, since the corresponding Pick matrix may not be positive definite. However, there is always an $\epsilon > 0$ such that f_λ exists whenever $1 - \epsilon < |\lambda| < 1$, hence the limit results are valid.

Proof. Let $M_\lambda = \{g : g \in \mathcal{S}, g(z_k) = \frac{w_k}{b_\lambda(z_k)}\}$. First note that if $g \in M_\lambda$, then gb_λ satisfies the interpolation conditions. Furthermore

$$\mathbb{K}_\Psi(g) = \mathbb{K}_\Psi(gb_\lambda) \geq \mathbb{K}_\Psi(f_\lambda) \geq \mathbb{K}_\Psi(\hat{f})$$

by the definitions of f_λ and $\hat{f}$. If we prove that

$$\min_{g \in M_\lambda} \mathbb{K}_\Psi(g) \to \mathbb{K}_\Psi(\hat{f}) \text{ as } |\lambda| \to 1, \tag{B.11}$$

then $\mathbb{K}_\Psi(f_\lambda) \to \mathbb{K}_\Psi(\hat{f})$, and by Theorem B.5 it follows that $f_\lambda \to \hat{f}$ in H_2. However, since $\frac{w_k}{b_\lambda(z_k)} \to w_k$ as $|\lambda| \to 1$, there is a sequence of functions $g_\lambda \in M_\lambda$ such that $g_\lambda \to \hat{f}$ in H_∞. By H_∞ continuity of $\mathbb{K}_\Psi$, (B.11) holds. □

Note that Theorem B.8 holds for any positive and continuous Ψ, whereas Theorem B.4 only holds if $\Psi = |\sigma|^2$ and σ belong to $\mathcal{K}_0$. This is because the proof of Theorem B.4 requires the use of Proposition B.7, where $\sigma \in \mathcal{K}_0$ is a condition.

We are now in a position to prove Theorem B.4. To this end, let $g_\lambda = b_w(f_\lambda)$ and $g = b_w(\hat{f})$. By Proposition B.7 and Theorem B.1, g_λ is the unique minimizer of $\mathbb{K}_\Psi(f)$ such that $f(z_k) = b_w(w_k)$, $k = 1, \ldots, n$, and $f(\lambda) = 0$. Furthermore g is the unique minimizer of $\mathbb{K}_\Psi(f)$ such that $f(z_k) = b_w(w_k)$, $k = 1, \ldots, n$. By Theorem B.8, $g_\lambda \to g$ in H_2. Since b_w is Lipschitz continuous, $f_\lambda \to \hat{f}$ in H_2. This concludes the proof of Theorem B.4.

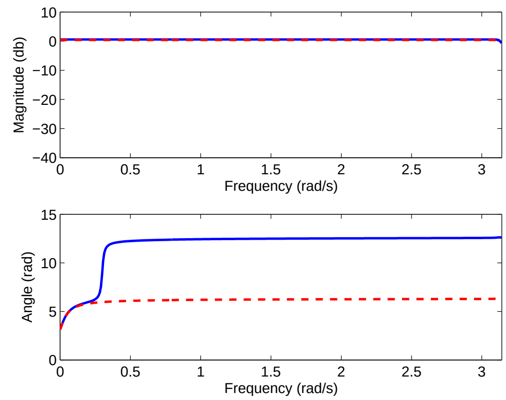

Figure B.5: *The sensitivity functions corresponding to the maximum entropy solution, i.e.* $\Psi \equiv 1$. *The dashed lines correspond to interpolation conditions* (B.7) *and* (B.8) *and the solid lines* (B.7)-(B.9).

Remark B.2. Consider a Pick problem with $z_0 = 0$ and several interpolation points close to the unit circle. Since $1 \in \mathcal{K}_0$ for all $n \geq 1$, it follows from Theorem B.4 thet removing the interpolation points close to the boundary have little effect on the minimium-entropy interpolant. Unless the interpolation value equals zero (Theorem B.8), the situation is more complicated for a more general choice of Ψ, since removing an interpolation condition generally produces a $\sigma \notin \mathcal{K}_0$.

B.5 Revisiting the example

We now return to the example of Section B.3. For determining the sensitivity function, two design tools were used, namely adding interpolation conditions and changing the weight Ψ by adding spectral zeros. We shall now investigate how these strategies have affected the design.

As a starting point we choose the maximum entropy interpolant corresponding to the interpolation conditions (B.7) and (B.8). This senitivity function, obtained by using the weight $\Psi \equiv 1$, is depicted Figure B.5 with a dashed line. Next we observe what happens to the maximum entropy solution when we add the interpolation conditions (B.9), which requires the interpolant to be zero at the points $0.9901e^{\pm 0.3i}$. This interpolation point is close to the unit circle. The corresponding sensitivity function is depicted by the solid line in Figure B.5. As is seen, the added interpolation points have neglible effect on the modulus of the interpolant and only a local effect on the phase around 0.3 rad/sec, where there is a sharp shift of 2π in the phase. This is in harmony with Theorem B.4 which states that the effect of adding additional interpolation points close to the unit circle is small in the H_2-norm.

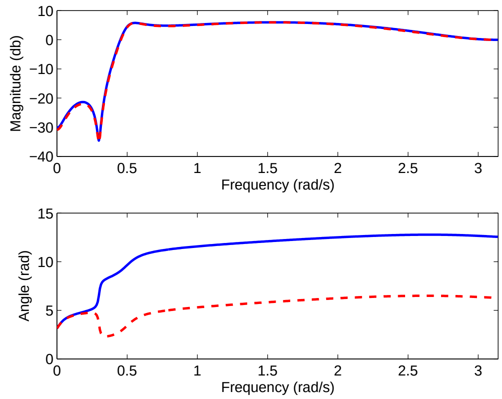

Figure B.6: *The sensitivity functions corresponding to the weight* Ψ_N. *The dashed lines correspond to interpolation conditions* (B.7) *and* (B.8) *and the solid lines to* (B.7)-(B.9).

However, the theory allows for shaping the interpolant by specifying spectral zeros or, equivalently, the weight Ψ. In the motivating example of Section B.3, the spectral zeros were chosen to be in $0.97e^{\pm 0.55i}$ and $0.9e^{\pm 1.55i}$, which corresponds to the weight Ψ_{N} given by (B.10). The sensitivity function S_{N} obtained, via (B.6), by using this weight and the interpolation constraints (B.7)-(B.9) is depicted in Figure B.6 with solid line. If we remove the interpolation condition (B.9) we obtain the sensitivity function depicted by dashed line in Figure B.6. As can be seen, also in this case, the only significant change resulting from the additional interpolation condition is the change of 2π in the phase around $\theta = 0.3$. As before this change is very local close to the added interpolation condition.

Therefore, solely adding the interpolation condition (B.9) does not change the solution significantly. The change in the magnitude is neglible, as is the change in the phase, except for the region close to the added interpolation point, where there is a sharp shift of 2π in the phase. Since the shift occurs over a short interval, the change in H_2 norm is minor, and as the interpolation point approaches the boundary this shift will have negligible effect on the H_2 norm. This example shows why the same convergence result could not hold for the H_∞ norm.

Let us return to Figure B.3 and the fact that there is no significant change in the sensitivity in the high frequency area, despite the large weight in this region. This could be due to the interpolation condition (B.8), which lies very close to the boundary in the high frequency region. Note that any effect (B.8) has on the interpolant is not in conflict with Theorem B.4. This is because $\sigma \notin \mathcal{K}_0$ if (B.8) is removed, and hence Theorem B.4 is not applicable. What can be concluded is that the weight has a large effect on the design. This is further exploited in [9] where a systematic procedure for finding appropriote weights is developed.

B.6 Proofs

For the proof of Theorem B.5 we will use concepts from convex analysis. Let $\Lambda : \mathbb{X} \to \mathbb{R}$ be a strictly convex functional, where $\mathbb{X}$ is compact and convex. Then the minimum

$$\beta = \min_{x \in \mathbb{X}} \Lambda(x)$$

exists and is attained at a unique $x \in \mathbb{X}$. Consider the set K_ϵ of $\epsilon-$suboptimal solutions

$$K_\epsilon = \{x \in \mathbb{X} : \Lambda(x) < \beta + \epsilon\},\ \epsilon > 0.$$

It seems reasonable that the "size" of K_ϵ tends to zero as $\epsilon \to 0$. However, to state and prove this properly, we need topological considerations and the concept of strong convexity.

Definition B.9. A functional Λ is *strongly convex with respect to the norm* $\|\cdot\|$ if there exists an $\alpha : [0, \infty) \to [0, \infty)$ that is continuous, strictly increasing and satisfies $\alpha(0) = 0$, for which

$$\frac{1}{2}(\Lambda(x) + \Lambda(y)) \geq \Lambda(\frac{x+y}{2}) + \alpha(\|x - y\|)$$

holds for all $x, y \in \mathbb{X}$.

Lemma B.10. *Let $\mathbb{X}$ be a convex set, and let Λ be a strongly convex functional on $\mathbb{X}$ with respect to the norm $\|\cdot\|$. Moreover, let $\hat{x}$ be the minimum of $\Lambda(x)$ such that $x \in \mathbb{X}$. Then $\Lambda(x_k) \to \Lambda(\hat{x}), x_k \in \mathbb{X}$, implies $\|x_k - \hat{x}\| \to 0$.*

An equivalent statment is that, if Λ is strongly convex with respect to the norm $\|\cdot\|$, then

$$\sup\{\|x - y\| : x, y \in K_\epsilon\} \to 0$$

as $\epsilon \to 0$, or, equivalently, K_ϵ is a neighboorhood basis for the optimal point $\hat{x}$ in the topology induced by the norm $\|\cdot\|$.

Proof. Assume that the statement of Lemma B.10 does not hold, i.e. that there exists an $\epsilon > 0$ so that for any $\delta > 0$ it is possible to find an $x \in \mathbb{X}$ so that $|\Lambda(x) - \Lambda(\hat{x})| < \delta$ and $\|x - \hat{x}\| > \epsilon$. Let $\delta < \alpha(\epsilon)$. Then $\Lambda(\hat{x}) + \delta \geq \frac{1}{2}(\Lambda(x) + \Lambda(\hat{x})) \geq \Lambda(\frac{x+\hat{x}}{2}) + \alpha(\|x - \hat{x}\|) \geq \Lambda(\frac{x+\hat{x}}{2}) + \alpha(\epsilon)$. This contradicts that $\hat{x}$ is the minimizer, and hence the validity of Lemma B.10 is proved by contradiction. □

In order to apply this result to the entropy functional, we need to show that $\mathbb{K}_\Psi$ is strongly convex with respect to the H_2 norm.

Proposition B.11. *Let $\Psi \in C(\mathbb{T})_+$. Then the entropy functional $\mathbb{K}_\Psi$ is strongly convex with respect to the H_2 norm.*

Proof. For $|f| < 1$ and $|g| < 1$, we have the following inequality

$$\tfrac{1}{2}(-\log(1-|f|^2) - \log(1-|g|^2)) \geq -\log\left(1 - \left|\tfrac{f+g}{2}\right|^2\right) + \tfrac{1}{2}\log\left(1 + \tfrac{|f-g|^2}{2}\right). \tag{B.12}$$

To see this, use the parallellogram law

$$|f+g|^2 + |f-g|^2 = 2|f|^2 + 2|g|^2$$

to obtain

$$\begin{aligned}\left(1 - \frac{|f+g|^2}{4}\right)^2 &= (1-|f|^2)(1-|g|^2) - |f|^2|g|^2 + \frac{|f-g|^2}{2} \\ &\quad + \frac{1}{16}\left(2|f|^2 + 2|g|^2 - |f-g|^2\right)^2 \\ &= (1-|f|^2)(1-|g|^2) + \frac{|f-g|^2}{4}\left(2 - |f|^2 - |g|^2\right) \\ &\quad + \frac{1}{4}\left(|f|^2 - |g|^2\right)^2 + \frac{1}{16}|f-g|^4 \\ &\geq (1-|f|^2)(1-|g|^2) + \frac{|f-g|^2}{4}\left(2 - |f|^2 - |g|^2\right).\end{aligned}$$

Consequently, since

$$\frac{2 - |f|^2 - |g|^2}{(1-|f|^2)(1-|g|^2)} \geq 2,$$

we have

$$\frac{\left(1 - \frac{|f+g|^2}{4}\right)^2}{(1-|f|^2)(1-|g|^2)} \geq 1 + \frac{|f-g|^2}{2},$$

from which (B.12) follows. Then, multiplying (B.12) by Ψ and integrating, we obtain

$$\frac{1}{2}(\mathbb{K}_\Psi(f) + \mathbb{K}_\Psi(g)) \geq \mathbb{K}_\Psi\left(\frac{f+g}{2}\right) + \frac{1}{2}\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f-g|^2}{2}\right) dm. \tag{B.13}$$

Since $\log(1+t) \geq t/2$ for $t \in [0,2]$, the last term in (B.13) is bounded from below by

$$\frac{1}{2}\int_{\mathbb{T}} \Psi \log\left(1 + \frac{|f-g|^2}{2}\right) dm \geq \int_{\mathbb{T}} \Psi \frac{|f-g|^2}{4} dm \geq \frac{\min \Psi}{4}\|f-g\|_2^2,$$

establishing the strong convexity of $\mathbb{K}_\Psi$ □

Theorem B.5 then follows from Lemma B.10 and Proposition B.11.

We now provide a more direct proof of Proposition B.7. In fact, $b_\lambda(f)$ clearly satisfies the interpolation conditions $b_\lambda(f(z_k)) = b_\lambda(w_k)$ and $\|b_\lambda(f)\|_\infty < 1$. Let $f = b/a$. Then

$$\frac{\beta}{\alpha} = b_\lambda(f) = \frac{\lambda - f}{1 - \bar{\lambda} f} = \frac{a\lambda - b}{a - \bar{\lambda} b},$$

and hence

$$\alpha\alpha^* - \beta\beta^* = (a - \bar{\lambda}b)(a - \bar{\lambda}b)^* - (a\lambda - b)(a\lambda - b)^* = (1 - \lambda\bar{\lambda})(aa^* - bb^*).$$

This shows that $g = b_\lambda(f)$ is the interpolant corresponding to σ.

B.7 Conclusions and further directions

In this paper we have studied the (generalized) entropy functional $\mathbb{K}_\Psi$ of [4] and the interpolants solving the optimization problem

$$\min \mathbb{K}_\Psi(f) \text{ subject to } f(z_k) = w_k,\ k = 0, 1, \ldots, n.$$

It is shown that, if the entropies of a sequence of interpolants converge to the minimum, then the corresponding interpolants converge in H_2, but not necessarily in H_∞. Furthermore, if the interpolation values are transformed by a Möbius transform, so is the minimizing interpolant. Next we show that the introduction of an additional interpolation point close to the boundary produces an insignificant increase in the entropy gain. Taken together with the results above, this implies that the change in the interpolant is small in H_2 norm.

We have analyzed a design example from robust control, studied by Nagamune [11], in the context of our results. The effect of an added interpolation condition close to the boundary turns out to be small, in harmony with Theorems B.4 and B.8. In the solution of Nagamune [11], a main objective of the additional interpolation conditions was to increase the dimension of $\mathcal{K}$, thereby allowing for more design parameters. However, by instead applying the parameterization of Theorem B.3, we could solve the optimization problem for a larger class of Ψ without the additional interpolation conditions. Adding interpolation points restricts the admissible set, and, if they have negligible effect, one would expect better solutions without them. A complete theory for this is developed in [9].

Acknowledgments

The authors would like to thank Professor Krister Svanberg for his input regarding the strong convexity.

B.8 Bibliography

[1] J. Agler and J.E. McCarthy. *Pick Interpolation and Hilbert Function Spaces*. American Mathematical Society, 2002.

[2] A.C. Antoulas. A new result on passivity preserving model reduction. *Systems and Control Letters*, 54:361–374, 2005.

[3] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[4] C. I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski. Generalized interpolation in H-infinity with a complexity constraint. *Transactions of the American Mathematical Society*, 358:965–987, 2006.

[5] C. I. Byrnes and A. Lindquist. On the duality between filtering and Nevanlinna-Pick interpolation. *SIAM J. Control and Optimization*, 39:757–775, 2000.

[6] C. I. Byrnes and A. Lindquist. The generalized moment problem with complexity constraint. *Integral Equations and Operator Theory*, 56(2):163–180, 2006.

[7] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory*. Macmillan Publishing Company, New York, 1992.

[8] G. Fanizza, J. Karlsson, A. Lindquist, and R. Nagamune. Passivity-preserving model reduction by analytic interpolation. *Linear Algebra and its Applications*, 425:608–633, 2007.

[9] J. Karlsson, T. Georgiou, and A Lindquist. The inverse problem of analytic interpolation with degree constraint and weight selection for control synthesis. *provisionally accepted to IEEE Transactions of Automatic Control*, 2008.

[10] D. Mustafa and K. Glover. *Minimum Entropy H-infinity Control*. Springer-Verlag, Berlin Heidelberg, 1990.

[11] R. Nagamune. Closed-loop shaping based on the Nevanlinna-Pick interpolation with a degree bound. *IEEE Trans. Automat. Control*, 49:300–305, 2004.

[12] D.C. Sorensen. Passivity preserving model reduction via interpolation of spectral zeros. *Systems and Control Letters*, 54:347–360, 2005.

Stability-Preserving Rational Approximation Subject to Interpolation Constraints

Johan Karlsson and Anders Lindquist

Abstract

A quite comprehensive theory of analytic interpolation with degree constraint, dealing with rational analytic interpolants with an a priori bound, has been developed in recent years. In this paper we consider the limit case when this bound is removed, and only stable interpolants with a prescribed maximum degree are sought. This leads to weighted H_2 minimization, where the interpolants are parameterized by the weights. The inverse problem of determining the weight given a desired interpolant profile is considered, and a rational approximation procedure based on the theory is proposed. This provides a tool for tuning the solution to specifications.

C.1 Introduction

Stability-preserving model reduction is a topic of major importance in systems and control, and over the last decades numerous such approximation procedures have been developed; see, e.g., [3, 17, 23, 1] and references therein. In this paper we introduce a novel approach to stability-preserving model reduction that also accommodates interpolation contraints, a requirement not uncommon in systems and control. By choosing the weights appropriately in a family of weighted H_2 minimization problems, the minimizer will both have low degree and match the original system.

As we shall see in this paper, stable interpolation with degree constraint can be regarded as a limit case of bounded analytic interpolation under the same degree constraint – a topic that has been thoroughly researched in recent years; see [5, 9]. More precisely, let f be a function in $H(\mathbb{D})$, the space of functions analytic in the unit disc $\mathbb{D} = \{z : |z| < 1\}$, satisfying

(i) the interpolation condition

$$f(z_k) = w_k, \quad k = 0, \dots, n, \tag{C.1}$$

(ii) the a priori bound $\|f\|_\infty \leq \gamma$, and

(iii) the condition that f be rational of degree at most n,

where $z_0, z_1, \dots, z_n \in \mathbb{D}$ are distinct (for simplicity) and $w_0, w_1, \dots, w_n \in \mathbb{C}$. It was shown in [5] that, for each such f, there is a rational function $\sigma(z)$ of the form

$$\sigma(z) = \frac{p(z)}{\tau(z)}, \qquad \tau(z) := \prod_{k=0}^{n} (1 - \bar{z}_k z),$$

where $p(z)$ is a polynomial of degree n with $p(0) > 0$ and $p(z) \neq 0$ for $z \in \mathbb{D}$ such that f is the unique minimizer of the generalized entropy functional

$$-\int_{-\pi}^{\pi} |\sigma(e^{i\theta})|^2 \gamma^2 \log(1 - \gamma^{-2}|f(e^{i\theta})|^2) \frac{d\theta}{2\pi}$$

subject to the interpolation conditions (C.1). In fact, there is a complete parameterization of the class of all interpolants satisfying (i)–(iii) in terms of the zeros of σ, which also are the spectral zeros of f; i.e., the zeros of $\gamma^2 - f(z)f^*(z)$ located in the complement of the unit disc. It can also be shown that this parameterization is smooth, in fact a diffeomorphism [6].

This smooth parameterization in terms of spectral zeros is the center piece in the theory of analytic interpolation with degree constraints; see [4, 5] and reference therein. By tuning the spectral zeros one can obtain an interpolant that better fulfills additional design specifications. However, one of the stumbling-blocks in the application of this theory has been the lack of a systematic procedure for achieving this tuning. In fact, the relation between the spectral zeros of f and f itself is nontrivial, and how to choose the spectral zeros in order to obtain an interpolant which satisfy the given design specifications has been a partly open problem.

In order to understand this problem better, we will in this paper focus on the limit case as $\gamma \to \infty$; i.e., the case when condition (ii) is removed. We shall refer to this problem – which is of considerable interest in its own right – as *stable interpolation with degree constraint*. Note that, as $\gamma \to \infty$,

$$-\gamma^2 \log(1 - \gamma^{-2}|f|^2) \to |f|^2,$$

and hence (see Proposition C.2),

$$-\int_{-\pi}^{\pi} |\sigma|^2 \gamma^2 \log(1 - \gamma^{-2}|f|^2) \frac{d\theta}{2\pi} \to \int_{-\pi}^{\pi} |\sigma f|^2 \frac{d\theta}{2\pi}.$$

For the case $\sigma \equiv 1$, this connection between the H_2 norm and the corresponding entropy functional have been studied in [15]. Consequently, the stable interpolants with degree constraint turn out to be minimizers of weighted H_2 norms. Indeed, the H_2 norm plays the same role in stable interpolation as the entropy functional does in bounded interpolation. Stable interpolation and H_2 norms are considerably easier to work with than bounded analytic interpolation and entropy functionals, but many of the concepts and ideas are similar.

The purpose of this paper is twofold. First, we want to provide a stability-preserving model reduction procedure that admits interpolation constraints and error bounds. Secondly, this theory is the simplest and most transparent gateway for understanding the full power of bounded analytic interpolation with degree constraint. In fact, our paper provides, together with the results in [12], the key to the problem of how to settle an important open question in the theory of bounded analytic interpolation with degree constraint, namely how to choose spectral zeros. In the present setting, spectral zeros are actually the poles.

In many applications, no interpolation conditions (or only a few) are given a priori. This allows us to use the interpolation points as additional tuning variables, available for satisfying design specifications. Such an approach for passivity-preserving model reduction was taken in [8]. However, a problem left open in [8] was how to actually select spectral zeros and interpolation points in a systematic way in order to obtain the best approximation. This problem, here in the context of stability-preserving model reduction, is one of the topics of this paper.

The paper is outlined as follows. In Section C.2 we show that the problem of stable interpolation is the limit, as the bound tend to infinity, of the bounded analytic interpolation problem stated above. In Section C.3 we derive the basic theory for how all stable interpolants with a degree bound may be obtained as weighted H_2-norm minimizers. In Section C.4 we consider the inverse problem of H_2 minimization, and in Section C.5 the inverse problem is used for model reduction of interpolants. The inverse problem and the model reduction procedure are closely related to the theory in [12]. A model reduction procedure where no a priori interpolation conditions are required are derived in Section C.6. This is motivated by a weighed relative error bound of the approximant and gives a systematic way to choose the interpolation points. This approximation procedure is also tunable so as to give small error in selected regions. In the Appendix we describe how the corresponding quasi-convex optimization problems can be solved. Finally, in Section C.7 we illustrate our new approximation procedures by applying them to an example and finally conclude with two control design examples.

C.2 Bounded interpolation and stable interpolation

In this section we show that the H_2 norm is the limit of a sequence of entropy functionals. From this limit, the relation between stable interpolation and bounded interpolation is established, and it is shown that some of the important concepts in the two different frameworks match.

First consider one of the main results of bounded interpolation: a complete parameterization of all interpolants with a degree bound [5]. For this, we will need two key concepts in that theory; the entropy functional

$$\mathbb{K}^{\gamma}_{|\sigma|^2}(f) = -\int_{-\pi}^{\pi} \gamma^2 |\sigma(e^{i\theta})|^2 \log(1-\gamma^{-2}|f(e^{i\theta})|^2)\frac{d\theta}{2\pi},$$

where we take $\mathbb{K}^{\gamma}_{|\sigma|^2}(f) := \infty$ whenever the H_∞ norm $\|f\|_\infty > \gamma$, and the co-invariant subspace

$$\mathcal{K} = \left\{ \frac{p(z)}{\tau(z)} \,:\, \tau(z) = \prod_{k=0}^{n}(1-\bar{z}_k z), p \in \mathrm{Pol}(n) \right\}. \tag{C.2}$$

Here $\mathrm{Pol}(n)$ denotes the set of polynomials of degree at most n, and $\{z_k\}_{k=0}^n$ are the interpolation points.

In fact, any interpolant f of degree at most n with $\|f\|_\infty \le \gamma$ is a minimizer of $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$ subject to (C.1) for some $\sigma \in \mathcal{K}_0$, where

$$\mathcal{K}_0 = \{\sigma \in \mathcal{K} \,:\, \sigma(0) > 0, \sigma \text{ outer}\}.$$

Furthermore, all such interpolants are parameterized by $\sigma \in \mathcal{K}_0$. This is one of the main results for bounded interpolation in [5] and is stated more precisely as follows.

Theorem C.1. *Let $\{z_k\}_{k=0}^n \subset \mathbb{D}, \{w_k\}_{k=0}^n \subset \mathbb{C}$, and $\gamma \in \mathbb{R}_+$. Suppose that the Pick matrix*

$$P = \left[\frac{\gamma^2 - w_k \bar{w}_\ell}{1 - z_k \bar{z}_\ell}\right]_{k,\ell=0}^{n} \tag{C.3}$$

is positive definite, and let σ be an arbitrary function in $\mathcal{K}_0$. Then there exists a unique pair of elements $(a,b) \in \mathcal{K}_0 \times \mathcal{K}$ such that

(i) $f(z) = b(z)/a(z) \in H^\infty$ *with* $\|f\|_\infty \le \gamma$

(ii) $f(z_k) = w_k, \quad k = 0,1,\dots,n$*, and*

(iii) $|a(z)|^2 - \gamma^{-2}|b(z)|^2 = |\sigma(z)|^2$ *for* $z \in \mathbb{T}$,

where $\mathbb{T} := \{z : |z| = 1\}$. Conversely, any pair $(a,b) \in \mathcal{K}_0 \times \mathcal{K}$ satisfying (i) and (ii) determines, via (iii), a unique $\sigma \in \mathcal{K}_0$. Moreover, the optimization problem

$$\min \mathbb{K}^{\gamma}_{|\sigma|^2}(f) \;\; \textit{s.t.} \;\; f(z_k) = w_k, \; k = 0,\dots,n$$

has a unique solution f that is precisely the unique f satisfying conditions (i), (ii) and (iii).

The essential content of this theorem is that the class of interpolants satisfying $\|f\|_\infty \le \gamma$ may be parameterized in terms of the zeros of σ, and that these zeros are the same as the *spectral zeros* of f; i.e., the zeros of the spectral outer factor $w(z)$ of $w(z)w^*(z) = \gamma^2 - f(z)f^*(z)$, where $f^*(z) = \overline{f(\bar{z}^{-1})}$.

Let $\|f\| = \sqrt{\langle f, f\rangle}$ denote the norm in the Hilbert space $H_2(\mathbb{D})$ with inner product

$$\langle f, g\rangle = \int_{-\pi}^{\pi} f(e^{i\theta})\overline{g(e^{i\theta})}\frac{d\theta}{2\pi}.$$

As the bound γ tend to infinity,

$$-\gamma^2 \log(1 - \gamma^{-2}|f|) \to |f|^2.$$

Therefore, the entropy functional $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$ converge to the weighted H_2 norm $\|\sigma f\|^2$.

Proposition C.2. *Let $f \in H_\infty(\mathbb{D})$ and σ be rational functions with σ outer. Then*

(i) $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$ is a non-increasing function of γ, and,

(ii) $\mathbb{K}^{\gamma}_{|\sigma|^2}(f) \to \|\sigma f\|^2$ as $\gamma \to \infty$.

Proof. It clearly suffices to consider only $\gamma \geq \|f\|_\infty$. Then the derivative of $-\gamma^2 \log(1 - \gamma^{-2}|f|^2)$ with respect to γ is non-positive for $|f| \leq \gamma$, and hence $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$ is non-increasing. To establish (ii), note that

$$-\gamma^2 \log(1 - \gamma^{-2}|f|^2) = |f|^2 + O(\gamma^{-2}|f|^2),$$

and therefore

$$-|\sigma|^2\gamma^2 \log(1 - \gamma^{-2}|f|^2) \to |\sigma f|^2$$

pointwise in $\mathbb{T}$ except for σ with poles in $\mathbb{T}$. There are two cases of importance. First, if σ has no poles in $\mathbb{T}$, or if a pole of σ coincided with a zero of f of at least the same multiplicity, then $-|\sigma|^2\gamma^2 \log(1 - \gamma^{-2}|f|^2)$ is bounded, and (ii) follows from bounded convergence. Secondly, if σ has a pole in $\mathbb{T}$ at a point in which f does not have a zero, then both $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$, and $\|\sigma f\|^2$ are infinite for any γ. □

The condition $\|f\|_\infty < \infty$ is needed in Proposition C.2. Otherwise, if $\|f\|_\infty = \infty$, then $\mathbb{K}^{\gamma}_{|\sigma|^2}(f)$ is infinite for any γ, while $\|\sigma f\|^2$ may be finite if σ has zeros in the poles of f on $\mathbb{T}$.

The next proposition shows that stable interpolation may be seen as the limit case of bounded interpolation when the bound γ tend to infinity.

Proposition C.3. *Let σ be any outer function such that the minimizer f of*

$$\min \|\sigma f\| \text{ subject to } f(z_k) = w_k, k = 0, \dots, n \tag{C.4}$$

satisfies $\|f\|_\infty < \infty$. Let f_γ be the minimizer of

$$\min \mathbb{K}^{\gamma}_{|\sigma|^2}(f_\gamma) \text{ subject to } f_\gamma(z_k) = w_k, k = 0, \dots, n$$

for $\gamma \in \mathbb{R}_+$ large enough so that the Pick matrix (C.3) *is positive definite. Then $\|\sigma(f - f_\gamma)\| \to 0$ as $\gamma \to \infty$.*

Proof. By Proposition C.2, and since f and f_γ are minimizers of the respective functional, we have

$$\mathbb{K}^{\gamma}_{|\sigma|^2}(f) \geq \mathbb{K}^{\gamma}_{|\sigma|^2}(f_\gamma) \geq \|\sigma f_\gamma\|^2 \geq \|\sigma f\|^2.$$

Moreover, since $\mathbb{K}^{\gamma}_{|\sigma|^2}(f) \to \|\sigma f\|^2$ as $\gamma \to \infty$ it follows that $\|\sigma f_\gamma\|^2 \to \|\sigma f\|^2$, and hence, by Lemma C.8, we have $\|\sigma(f - f_\gamma)\| \to 0$ as $\gamma \to \infty$, as claimed. □

Note that Proposition C.3 holds for any σ which is outer and not only for $\sigma \in \mathcal{K}_0$. However, if $\sigma \in \mathcal{K}_0$, then $\deg f_\gamma \leq n$ for any γ. Therefore, since $\|\sigma(f - f_\gamma)\| \to 0$ as $\gamma \to \infty$, for $\sigma \in \mathcal{K}_0$ the minimizer f of (C.4) will be a stable interpolant of degree at most n. We will return to this in the next section.

It is interesting to note how concepts in the two types of interpolation are related. First of all, the weighted H_2 norm plays the same role in stable interpolation as the entropy functional does in bounded interpolation. Secondly, the spectral zeros, which play a major role in degree constrained bounded interpolation, simply correspond to the poles in stable interpolation. This may be seen from (iii) in Theorem C.1.

C.3 Rational interpolation and H_2 minimization

In the previous section we have seen that minimizers of a specific class of H_2 norms are stable interpolants of degree at most n. This, and also the fact that this class may be parameterized by $\sigma \in \mathcal{K}_0$ can be proved using basic Hilbert space concepts. This will be done in this section.

To this end, first consider the minimization problem

$$\min \|f\| \text{ subject to } f(z_k) = w_k, k = 0, \dots, n, \tag{C.5}$$

without any weight σ. Let $f_0 \in H_2(\mathbb{D})$ satisfy the interpolation condition (C.1). Then any $f \in H_2(\mathbb{D})$ satisfying (C.1) can be written as $f = f_0 + v$, where $B = \prod_{k=0}^{n} \frac{z_k - z}{1 - \bar{z}_k z}$ and $v \in BH_2$. Therefore, (C.5) is equivalent to

$$\min_{v \in BH_2} \|f_0 + v\|.$$

By the Projection Theorem (see, e.g., [13]), there exists a unique solution $f = f_0 + v$ to this optimization problem, which is orthogonal to BH_2, i.e. $f \in \mathcal{K} := H_2 \ominus BH_2$.

Conversely, if $f \in \mathcal{K}$ and $f(z_k) = w_k$, for $k = 0, \dots, n$, then f is the unique solution of (C.5). To see this, note that any interpolant in $H_2(\mathbb{D})$ may be written as $f + v$ where $v \in BH_2$. However, since $v \in BH_2 \perp \mathcal{K} \ni f$, we have $\|f + v\|^2 = \|f\|^2 + \|v\|^2$, and hence the minimizer is f, obtained by setting $v = 0$.

We summarize this in the following proposition.

Proposition C.4. *The unique minimizer of* (C.5) *belongs to* $\mathcal{K}$. *Conversely, if* $f \in \mathcal{K}$ *and* $f(z_k) = w_k$, *for* $k = 0, \dots, n$, *then* f *is the minimizer of* (C.5).

Consequently, in view of (C.2), f is a rational function with its poles fixed in the mirror images (with respect to the unit circle) of the interpolation points. By introducing weighted norms, any interpolant with poles in prespecified points may be constructed in a similar way. In fact, the set of interpolants f of degree $\leq n$ may be parameterized in this way. One way to see this is by considering

$$\min \|\sigma f\| \text{ subject to } f(z_k) = w_k, k = 0, \dots, n, \tag{C.6}$$

where $\sigma \in \mathcal{K}_0$. Since σ is invertible in $H(\mathbb{D})$, (C.6) is equivalent to

$$\min \|\sigma f\| \text{ subject to } (\sigma f)(z_k) = \sigma(z_k) w_k, k = 0, \ldots, n.$$

According to Proposition C.4, this has the optimal solution $\sigma f = b \in \mathcal{K}$, and hence the solution of (C.6), $f = \frac{b}{\sigma}$, is rational of degree at most n. To see that any solution of degree at most n can be obtained in this way, note that any such interpolant f is of the form $f = \frac{b}{\sigma}, b \in \mathcal{K}, \sigma \in \mathcal{K}_0$. Since $\sigma f = b \in \mathcal{K}$ holds together with the interpolation condition (C.1) if and only if $\sigma(z_k) f(z_k) = \sigma(z_k) w_k$ for $k = 0, \ldots, n$, f is the unique solution of (C.6), by Proposition C.4. This proves the following theorem.

Theorem C.5. *Let $\sigma \in \mathcal{K}_0$. Then the unique minimizer of*

$$\min \|\sigma f\| \textit{ subject to } f(z_k) = w_k,\ k = 0, \ldots, n, \tag{C.7}$$

belongs to $H(\mathbb{D})$ and is rational of a degree at most n. More precisely,

$$f = \frac{b}{\sigma} \tag{C.8}$$

where $b \in \mathcal{K}$ is the unique solution of the linear system of equations

$$b(z_k) = \sigma(z_k) w_k, \quad k = 0, 1, \ldots, n. \tag{C.9}$$

Conversely, if f satisfies (C.8) *for some $b \in \mathcal{K}$ and the interpolation condition* (C.1) *holds, then f is the unique minimizer of* (C.7)*.*

In other words, the set of interpolants in $H(\mathbb{D})$ of degree at most n may be parameterized in terms of weights $\sigma \in \mathcal{K}_0$. Another way to look at this is that the poles of the minimizer (C.8) are specified by the zeros of σ and that the numerator $b = \beta/\tau$ is determined from the interpolation condition by solving the linear system of equations

$$\beta(z_k) = \tau(z_k)\sigma(z_k) w_k, \quad k = 0, 1, \ldots, n \tag{C.10}$$

for the $n+1$ coefficients $\beta_0, \beta_1, \ldots, \beta_n$ of the polynomial $\beta(z)$. This is a Vandermonde system have a unique solution (as long as the interpolation points $z_o, z_1, \ldots, z_n$ are distinct as here).

Note that this parameterization is not necessarily injective. If, for example, $w_k = 1$ for $k = 0, \ldots, n$, then there is a unique function f of degree at most n that satisfies $f(z_k) = w_k, k = 0, \ldots, n$. No matter how $\sigma \in \mathcal{K}_0$ is chosen, $b = \sigma$, and hence the minimizer of (C.6) will be $f \equiv 1$.

C.4 The inverse problem

In [12] we considered the *inverse problem of analytic interpolation*; i.e., the problem of choosing an entropy functional whose unique minimizer is a prespecified interpolant. In this section we will consider the counterpart of this problem for stable interpolation. To this end, let us first introduce the subclass $H_{\mathbb{Q}}(\mathbb{D})$ of log-integrable analytic functions in $H(\mathbb{D})$ for which the inner part is rational. In particular, the class of rational analytic functions belong to $H_{\mathbb{Q}}(\mathbb{D})$.

Suppose $f \in H_{\mathbb{Q}}(\mathbb{D})$ satisfies the interpolation condition (C.1). Then, when does there exist σ which is outer such that f is the minimizer of

$$\min \|\sigma f\| \text{ subject to } f(z_k) = w_k, k = 0, \dots, n?$$

We refer to this as the *inverse problem of H_2 minimization*, and its solution is given in the following theorem.

Theorem C.6. *Let $f \in H_{\mathbb{Q}}(\mathbb{D})$ satisfy the interpolation condition $f(z_k) = w_k, k = 0, \dots, n$. Then f is the minimizer of*

$$\min \|\sigma f\| \textit{ subject to } f(z_k) = w_k,\ k = 0, \dots, n, \tag{C.11}$$

where σ is outer if and only if $\sigma f \in \mathcal{K}$, in which case the minimizer is unique. Such a σ exists if and only if f has no more than n zeros in $\mathbb{D}$.

Proof. The function f is the minimizer of (C.11) if and only if $b = \sigma f$ is the minimizer (necessarily unique) of

$$\min \|b\| \text{ subject to } b(z_k) := w_k \sigma(z_k),\ k = 0, \dots, n,$$

which, by Proposition C.4, holds if and only if $\sigma f = b \in \mathcal{K}$. Such a σ only exists if f has less or equal to n zeros inside $\mathbb{D}$. To see this, first note that, if f has more than n zeros in $\mathbb{D}$, then σf has more than n zeros in $\mathbb{D}$ and can therefore not be of the form p/τ with $p \in \text{Pol}(n)$. On the other hand, if f has less or equal to n zeros in $\mathbb{D}$, then let $p = \prod(z - p_k)$ where p_k are the zeros of f, and set $\sigma := \frac{p}{f\tau}$. Then σ is outer and satisfies $\sigma f \in \mathcal{K}$. □

Theorem C.6 defines a map φ that sends σ to the unique minimizer f of the optimization problem (C.11); i.e.,

$$\sigma \mapsto f = \varphi(\sigma). \tag{C.12}$$

Let W_f denote the set of weights σ that give f as a minimizer of (C.11); i.e., the inverse image $\varphi^{-1}(f)$ of f. By Theorem C.6,

$$\begin{aligned} W_f &:= \varphi^{-1}(f) = \{\sigma \text{ outer } : \sigma f \in \mathcal{K}\} \\ &= \left\{\sigma = \frac{p}{f\tau} : p \in \text{Pol}(n) \setminus \{0\}, \frac{p}{f} \text{ outer}\right\}, \end{aligned} \tag{C.13}$$

i.e., W_f may be parameterized in terms of the polynomials $p \in \text{Pol}(n)$. For the condition that pf^{-1} is outer to hold for some $p \in \text{Pol}(n)$, it is necessary that f has at most n zeros in $\mathbb{D}$. This is in accordance with Theorem C.6. It is interesting to note that the dimension of W_f depends on the number of zeros of f inside $\mathbb{D}$. The more zeros f has inside $\mathbb{D}$, the more restricted is the class W_f. One extreme case is when f has no zeros inside $\mathbb{D}$. Then p could be any stable polynomial of degree n. The other extreme is when f has n zeros in $\mathbb{D}$, in which case p is uniquely determined up to a multiplicative constant.

C.5 Rational approximation with interpolation constraints

In this section the solution of the inverse problem (Theorem C.6) will be used to develop an approximation procedure for interpolants. Let $f \in H_{\mathbb{Q}}(\mathbb{D})$ be a function satisfying the interpolation condition (C.1). We want to construct another function $g \in H_{\mathbb{Q}}(\mathbb{D})$ of degree at most n satisfying the same interpolation condition such that g is as close as possible to f.

Let $\sigma \in W_f$; i.e., let σ be a weight and such that f is the minimizer of (C.11), and let ρ be close to σ. Then it is reasonable that the minimizer g of the optimization problem

$$\min \|\rho g\| \text{ subject to } g(z_k) = w_k,\ k = 0, \dots, n, \tag{C.14}$$

is close to f. This is the statement of the following theorem.

Theorem C.7. *Let $f \in H_{\mathbb{Q}}(\mathbb{D})$ satisfy the interpolation condition $f(z_k) = w_k, k = 0, \dots, n$, and let $\sigma \in W_f$. Moreover, let ρ be an outer function such that*

$$\left\| 1 - \left| \frac{\rho}{\sigma} \right|^2 \right\|_\infty = \epsilon, \tag{C.15}$$

and let g be the corresponding minimizer of (C.14). *Then*

$$\|\sigma(f-g)\|^2 \leq \frac{4\epsilon}{1-\epsilon} \|\sigma f\|^2. \tag{C.16}$$

For the proof we need the following lemma.

Lemma C.8. *Let let $g \in H_{\mathbb{Q}}(\mathbb{D})$ satisfy $g(z_k) = w_k$ for $k = 0, \dots, n$, and let f be the minimizer of* (C.11). *Then, if*

$$\|\sigma g\|^2 \leq (1+\delta)\|\sigma f\|^2,$$

we have

$$\|\sigma(f-g)\|^2 \leq 2\delta\|\sigma f\|^2.$$

Proof. From the parallelogram law we have,

$$\frac{1}{2}\left(\|\sigma f\|^2 + \|\sigma g\|^2\right) = \left\|\sigma \frac{f+g}{2}\right\|^2 + \left\|\sigma \frac{f-g}{2}\right\|^2.$$

Therefore, since f is the minimizer of (C.11), and hence $\|\sigma f\| \leq \|\sigma(f+g)/2\|$, it follows that

$$\|\sigma(f-g)\|^2 \leq 2(\|\sigma g\|^2 - \|\sigma f\|^2) \leq 2\delta\|\sigma f\|^2,$$

which concludes the proof of the lemma. □

Proof of Theorem C.7: In view of (C.15) we have

$$(1-\epsilon)|\sigma(e^{i\theta})|^2 \leq |\rho(e^{i\theta})|^2 \leq (1+\epsilon)|\sigma(e^{i\theta})|^2$$

for all $\theta \in [-\pi, \pi]$. Therefore, since g is the minimizer of (C.14), by (C.15), we have

$$\|\sigma g\|^2 \leq \frac{1}{1-\epsilon}\|\rho g\|^2 \leq \frac{1}{1-\epsilon}\|\rho f\|^2 \leq \frac{1+\epsilon}{1-\epsilon}\|\sigma f\|^2 = (1+\delta)\|\sigma f\|^2,$$

where $\delta := 2\epsilon/(1-\epsilon)$. Consequently (C.16) follows from Lemma C.8.

We have thus shown that if $\left|\frac{\rho(z)}{\sigma(z)}\right|$ is close to 1 for $z \in \mathbb{T}$, then $\|\sigma(f-g)\|$ is small. This suggests the following approximation procedure, illustrated in Figure C.1. By Theorem C.5, the function φ, defined by (C.12), maps the subset $\mathcal{K}_0$ into the space of interpolants of degree at most n. In Figure C.1 these subsets are depicted by fat lines. The basic idea is to replace the hard problem of approximating f by a function g of degree at most n by the simpler problem of approximating an outer function σ by a function $\rho \in \mathcal{K}_0$.

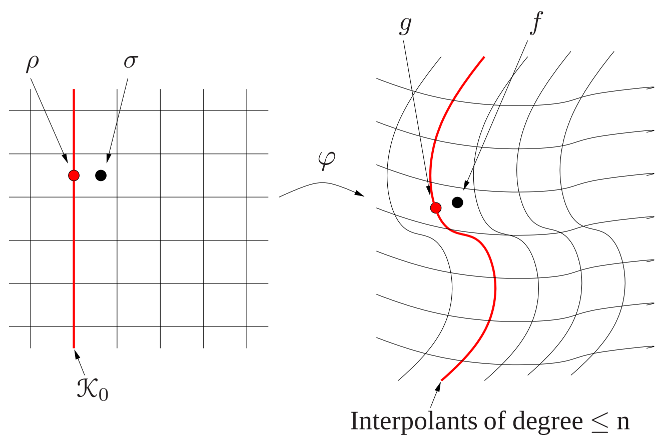

Figure C.1: *The map φ sending weighting functions to interpolants.*

Theorem C.7 suggests various strategies for choosing the functions $\rho \in \mathcal{K}_0$ and $\sigma \in W_f$ depending on the design preferences. If a small error bound for $\|\sigma(f-g)\|$ is desired for a particular $\sigma \in W_f$, this σ should be used together with the $\rho \in \mathcal{K}_0$ that minimizes (C.15).

However, obtaining a small value of (C.15) is often more important than the choice of σ. Therefore, in general it is more natural to choose the pair $(\sigma, \rho) \in (W_f, \mathcal{K}_0)$ that minimizes ϵ. For such a pair, setting $q := \tau\rho$, we can be see from (C.2) and (C.13) that

$$\epsilon = \left\| 1 - \left|\frac{\rho}{\sigma}\right|^2 \right\|_\infty = \left\| 1 - \left|\frac{qf}{p}\right|^2 \right\|_\infty, \tag{C.17}$$

where $q \in \text{Pol}(n)$ and $p \in \text{Pol}(n) \setminus \{0\}$ needs to be chosen so that p/f is outer. It is interesting to note that (C.17) is independent of $\tau(z) := \prod_{k=0}^{n}(1-\bar{z}_k z)$ and hence of the interpolation points $z_0, z_1, \ldots, z_n$.

Now suppose that f has ν zeros in $\mathbb{D}$; i.e., ν nonminimum-phase zeros. Then $f = \pi f_0$, where f_0 is outer (minimum phase) and π is an unstable polynomial of degree $\nu \leq n$. Setting $p = \pi p_0$, our optimization problem to minimize ϵ reduces to the problem to find a pair $(p_0, q) \in \text{Pol}(n-\nu) \times \text{Pol}(n)$ that minimizes

$$\epsilon = \left\| 1 - \left| \frac{q f_0}{p_0} \right|^2 \right\|_\infty \tag{C.18}$$

for a given nonminimum-phase f_0. This is a quasi-convex optimization problem, which can be solved as described in the Appendix (see also [20, 21]). The optimal q yields the optimal $\rho = q/\tau$. The approximant g is then obtained by solving the optimization problem (C.14) as described in Theorem C.5.

One should note that, the more zeros f has inside $\mathbb{D}$, the smaller is the choice of p. Therefore one expects approximations of non-minimum phase plants to be worse than approximations of plants without unstable zeros.

C.6 Rational approximation

In applications where there are no a priori interpolation constraints, the choice of interpolation points serve as additional design parameters. It is then important to choose them so that a good approximation is obtained. The main strategy previously used is to chose interpolation points close to the regions of the unit circle where good fit is desired. The closer to the unit circle the points are placed, the better fit, but the smaller is the region where good fit is ensured; see [8] for further discussions on this. However, in this paper we shall provide a systematic procedure for choosing the interpolation points, based on quasi-convex optimization.

As we have seen in the previous section the choice of interpolation points does not affect ϵ given by (C.17). However, since $\sigma = \frac{p}{f\tau}$, the weighted H_2 error bound (C.16) in Theorem C.7 becomes

$$\left\| \frac{p}{\tau} \frac{f-g}{f} \right\|^2 \leq \frac{4\epsilon}{1-\epsilon} \left\| \frac{p}{\tau} \right\|^2,$$

which depends on τ and hence on the choice of interpolation points. In fact, this is a weighed H_2 bound on the relative error $(f-g)/f$. If a specific part of the unit circle is of particular interest, interpolation points may be placed close to that part, which gives a bound on the weighted relative error with high emphasis on that specific region. (For a method to do this by convex optimization, see Remark C.3 in the next section.) If no particular part is more important than the rest, we suggest to select τ as the outer part of p; i.e., $|\tau(z)| = |p(z)|$ for $z \in \mathbb{T}$. This gives a natural choice of interpolation points that are the mirror images of the roots of τ. Furthermore, this choice gives the relative error bound $\|(f-g)/f\| \leq 4\epsilon/(1-\epsilon)$. This is summarized in the following theorem.

Theorem C.9. *Let p and q be polynomials of degrees at most n such that pf^{-1} is outer, and set*

$$\epsilon := \left\| 1 - \left| \frac{qf}{p} \right|^2 \right\|_\infty . \tag{C.19}$$

Let $z_0, z_1, \dots, z_n \in \mathbb{D}$ and let

$$g = \arg\min \|\rho g\| \text{ subject to } g(z_k) = f(z_k), k = 0, \dots, n,$$

where $\rho = q/\tau$ and $\tau = \prod_{k=0}^{n}(1 - \bar{z}_k z)$. Then $\deg g \leq n$ and

$$\left\| \frac{p}{\tau} \frac{f-g}{f} \right\|^2 \leq \frac{4\epsilon}{1-\epsilon} \left\| \frac{p}{\tau} \right\|^2 . \tag{C.20}$$

In particular, if the interpolation points $z_0, z_1, \dots, z_n$ are chosen so that $|\tau(z)| = |p(z)|$ for $z \in \mathbb{T}$, then

$$\left\| \frac{f-g}{f} \right\|^2 \leq \frac{4\epsilon}{1-\epsilon}. \tag{C.21}$$

Remark C.1. Note that the choice $|\tau| = |p|$ in Theorem C.9 implies that the unstable zeros of f become interpolation points. Therefore, for $\epsilon < 1$, $(f-g)/f$ belongs to H_2.

Remark C.2. Our method requires that we choose n to be greater than or equal to the number of unstable zeros. This is a natural design restriction, since the approximation problem becomes more difficult the larger is the number of unstable zeros. It should be noted that other methods for which there is a bound for the relative error, such as balanced stochastic truncation or Glover's relative error method (see, e.g., [10]), will not work either if the number of unstable zeros exceeds n. In fact, in such a case the corresponding phase function will have more than n Hankel singular values equal to 1, and therefore the bound will be infinite, and the problem to minimize $\|(f-g)/f\|_\infty$ over all g of degree at most n will have the optimal solution $g \equiv 0$. Also note that, unlike these methods, our method does not require f to be rational.

C.7 The computational procedure and some illustrative examples

Next we summarize the computational procedure suggested by the theory presented above and apply it to some examples.

Given a function $f \in H_{\mathbb{Q}}(\mathbb{D})$ with at most n zeros in $\mathbb{D}$, we want to construct a function $g \in H_{\mathbb{Q}}(\mathbb{D})$ of degree at most n that approximates f as closely as possible. We consider two versions of this problem. First we assume that f satisfies the interpolation condition (C.1), and we require g to satisfy the same interpolation conditions. Secondly, we relax the problem by removing the interpolation constraints.

Suppose that f has $\nu \leq n$ zeros in $\mathbb{D}$. Then $f = \pi f_0$, where f_0 is minimum-phase, and π is a polynomial of degree ν with zeros in $\mathbb{D}$. The approximant g can then be determined in two steps:

(i) Solve the quasi-convex optimization problem to find a pair $(p_0, q) \in \text{Pol}(n-\nu) \times \text{Pol}(n)$ that minimizes (C.18), as outlined in the Appendix. This yields optimal ϵ, p_0 and q. Set $p := \pi p_0$.

(ii) Solve the optimization problem (C.14) with $\rho = q/\tau$, as described in Theorem C.5. Exchanging σ for ρ in (C.10) we solve the Vandermonde system

$$\beta(z_k) = q(z_k)w_k, \quad k = 0, 1, \ldots, n,$$

for the $\beta \in \text{Pol}(n)$, which yields

$$g = \frac{\beta}{q} \tag{C.22}$$

and the bound (C.20), where $\tau(z) := \prod_{k=0}^{n}(1 - \bar{z}_k z)$.

For the problem without interpolation condition, before step (ii) we chose the interpolation points in one of the following ways.

(ii)$'$ Choose $z_0, z_1, \ldots, z_n$ arbitrarily, or as in Remark C.3 below. This yields a solution (C.22) and a bound (C.20).

(ii)$''$ Choose $z_0, z_1, \ldots, z_n$ so that τ is the outer (minimum-phase) factor of p. This yields a solution (C.22) and the bound (C.21) for the relative H_2 error.

Remark C.3. If a bound on the weighted error $\|w(f-g)\|$ is desired in Step (ii)$'$, it is natural to choose τ (which specify the interpolation points) so that $\frac{p}{\tau f}$ is as close to w as possible. This may be done by solving the convex optimization problem to find a $\tau \in \text{Pol}(n)$ that minimizes

$$\left\| 1 - \left| \frac{\tau f w}{p} \right|^2 \right\|_\infty,$$

as in the Appendix. If instead we need a bound on the weighted relative error $\|w(f-g)/f\|$, we modify the optimization problem accordingly.

Next, we apply these procedures to some numerical examples.

C.7.1 Model reduction with and without interpolation constraints

Example C.1. Let

$$f(z) = \frac{b(z)}{a(z)}$$

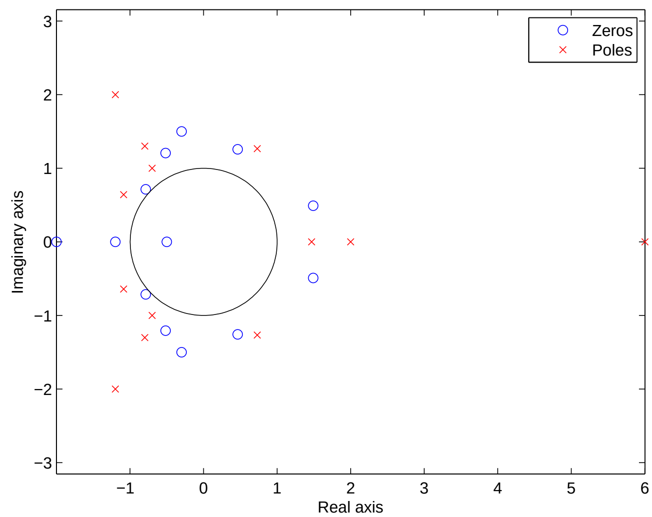

Figure C.2: *Poles and zeros of f in Examples 1 and 2.*

be the stable system of order 13 given by

$$\begin{aligned} b(z) &= 30z^{13} + 90z^{12} + 128.6z^{11} + 114.6z^{10} \\ &\quad -137.4z^9 - 322.3z^8 - 371.4z^7 + 10.8z^6 \\ &\quad +1005.8z^5 + 2428.7z^4 + 3967.0z^3 + 4189.7z^2 \\ &\quad +2800.6z + 726.2, \\ a(z) &= 4.0z^{13} - 13.4z^{12} - 44.2z^{11} - 144.5z^{10} \\ &\quad +83.5z^9 + 363.7z^8 + 791.4z^7 + 340.1z^6 \\ &\quad +770.7z^5 + 877.3z^4 - 93.6z^3 - 4767.8z^2 \\ &\quad -6349.3z - 4532.7. \end{aligned}$$

This system has one minimum-phase zero. The poles and zeros are given in Figure C.2.

Consider the problem to approximate f by a function g of degree six while preserving the values in the points $(z_0, z_1, \ldots, z_n) = (0, 0.3, 0.5, -0.1, -0.7, -0.3 \pm 0.3i)$. Such an interpolation condition occurs in certain applications.

Step (i) to solve the quasi-convex optimization problem to minimize (C.18) yields optimal ϵ, p and q, and Step (ii) the approximant g, the Bode plot of which is depicted in Figure C.3 together with that of f. The third subplot in the picture shows the relative error

$$\left| \frac{f(e^{i\theta}) - g(e^{i\theta})}{f(e^{i\theta})} \right| \text{ for } \theta \in [0, \pi].$$

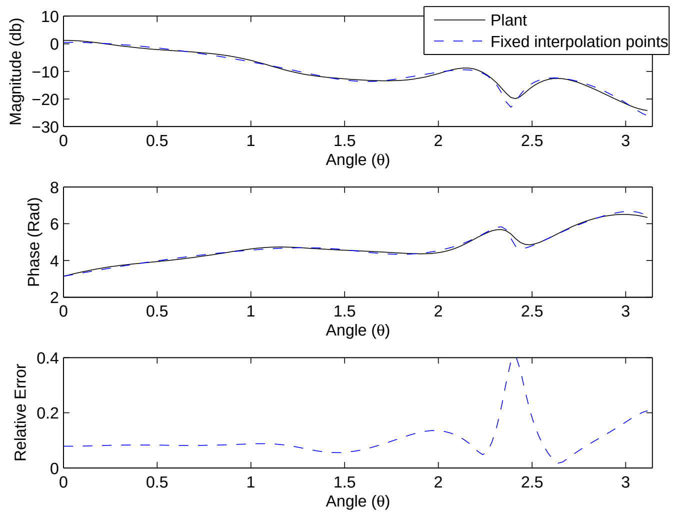

Figure C.3: *Bode plots of f and g together with the relative error.*

It is important to note that the function g, which is guaranteed to be stable, satisfies the prespecified interpolation conditions and the error bound (C.20). Figure C.3 shows that g matches f quite well.

Example C.2. Next we approximate the function f in Example 1 without imposing any interpolation condition. For $n = 4$, 6 and 8 we determine an approximant g_n of degree n via Steps (i) and (ii)$''$. This approximant satisfies the relative error bound (C.21). Then we compare g_n to an approximant $\hat{f}_n$ of the same degree obtained by balanced truncation [19, 24].

Since balance truncation imposes a bound on the absolute, rather than the relative, error, it is reasonable to also compare it with the approximant h_n of degree n obtained by stochastically balanced truncation [22, 18], which comes with a relative error bound.

The respective Bode plots and relative errors for the three methods are depicted in Figures C.4, C.5, and C.6. Stochastically balanced truncation gives the best approximation close to the valleys of the plant, and balanced truncation gives best approximation close to the peaks. The proposed method performs somewhere in between and has a more uniform relative error. In fact, as can be seen from Figure C.5 and Figure C.6, it is the method with the smallest relative L_∞-error for $n = 6$ and $n = 8$. As can be seen in the following tables, listing the relative and absolute errors of the three methods, the approximants of roughly the same quality.

Relative L_2 Error		Degree	
Approximation method	4	6	8
Proposed method	0.4736	0.0764	0.0194
Balanced truncation	0.4727	0.0785	0.0220
Stoch. Bal. truncation	0.7958	0.0656	0.0334

H_2 Error		Degree	
Approximation method	4	6	8
Proposed method	0.1918	0.0422	0.0100
Balanced truncation	0.0746	0.0451	0.0057
Stoch. Bal. truncation	0.3073	0.0506	0.0213

In the present example, the error bound (C.21) is quite conservative. In fact, the bound is $10.4735, 0.8765$, and 0.3994, for n equal to $4, 6$, and 8 respectively, which should be compared with the corresponding errors in the table. By comparison, the relative L_∞ bound on h_n is $3.9288, 0.3562$, and 0.0573 for n equal to $4, 6$, and 8 respectively, which is also conservative for $n = 4, 6$. Although these bounds are measured in different norms, it is still interesting to compare them. How to improve our bound will be subject to further studies.

In Figure C.7 the approximant g from Example 1 is compared to g_6. The interpolation points for g_6 are chosen according to (ii)$''$, and the interpolation condition of g is prespecified. It can be seen from Figure C.7 that g_6 matches f better than does g. This is because the interpolation points could be chosen freely for g_6.

Remark C.4. Note that stable approximation could be done iteratively by solving a nonconvex optimization directly by gradient methods to find local optima; see e.g. [14] and references therein. With a sufficiently good starting point, a global optimum could be obtained. Using such methods for the present example, it is possible to find approximants of degrees 4, 6, and 8 with relative errors that compare favorable to all the above methods. However, our method is based on convex and quasi-convex optimization, and thus does not rely on a good starting point, which is often difficult to find. It will be subject to further research to investigate in which way optimal approximations of weights relate to optimal approximations of interpolants.

C.7.2 Sensitivity shaping in robust control

Finally, we apply our approximation procedure to loop-shaping by low-degree controllers in robust control, where interpolation conditions are needed to ensure internal stability. Given a plant P, a controller is often designed by shaping the sensitivity function

$$S = \frac{1}{1 - PC},$$

where P and C are the transfer functions of the plant and the controller respectively. In fact, the design specifications may often be translated into conditions on the sensitivity

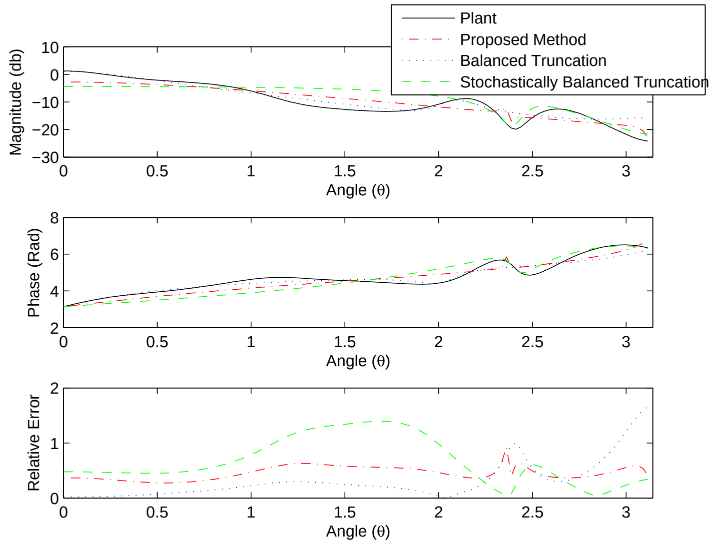

Figure C.4: *Bode plot of f, g_4, $\hat{f}_4$, and h_4 together with the relative errors.*

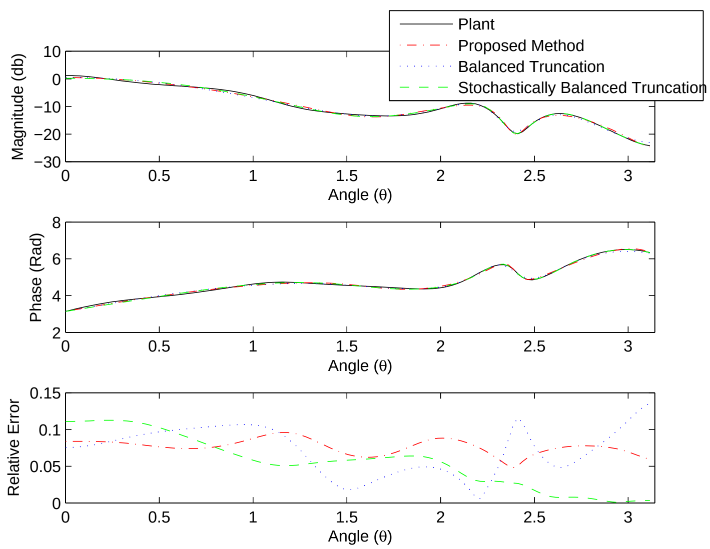

Figure C.5: *Bode plot of f, g_6, $\hat{f}_6$, and h_6 together with the relative errors.*

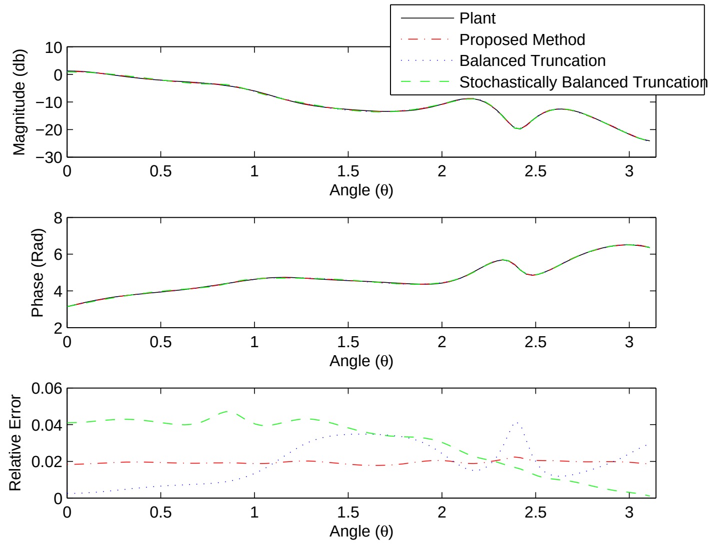

Figure C.6: *Bode plot of f, g_8, $\hat{f}_8$, and h_8 together with the relative errors.*

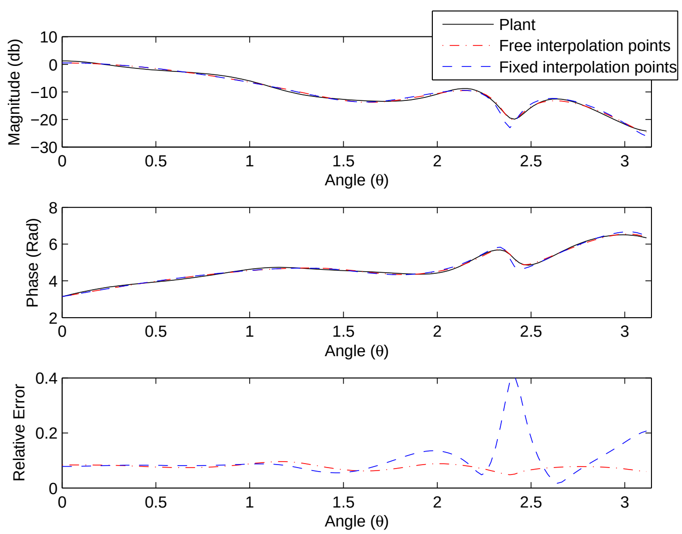

Figure C.7: *Bode plot of f, g_6, and g together with the relative errors.*

function. For internal stability of the closed loop system, the sensitivity function S needs to satisfy the following properties:

(i) S is analytic in $\mathbb{C}_+$ (continuous time) or in $\mathbb{D}^c$, where $\mathbb{D}^c := \{z \mid |z| \geq 1\}$, (discrete time),

(ii) $S(z_k) = 1$ whenever z_k is an unstable zero of P,

(iii) $S(p_k) = 0$ whenever p_k is an unstable pole of P.

Furthermore, in general we require that

(iv) S has low degree, and

(v) S satisfies additional design specifications.

The degree bound on S is important for several reasons. In fact, a low-degree sensitivity function results in a low-degree controller (see e.g., [16]), and, in some applications, the degree of the sensitivity function is important in its own right. A case in point is an autopilot, for which the feedback system itself is to be controlled. Conditions (i)-(iv) do not in general uniquely specify S, so the additional freedom can be utilized to satisfy additional design specifications (v).

Example C.3. As a simple example, also illustrating rational approximation of a nonrational function, consider sensitivity shaping of a feedback system with the plant $P(z) = (z-2)^{-1}$. Since P has an unstable pole at $z = 2$ and an unstable zero at $z = \infty$, the sensitivity function must satisfy $S(\infty) = 1$ and $S(2) = 0$. Then the function $f(z) := S(z^{-1})$ is analytic in $\mathbb{D}$, and satisfies

$$f(0) = 1 \text{ and } f(1/2) = 0. \tag{C.23}$$

Now, suppose we want to find a rational sensitivity function S_n of degree n that preserves internal stability and that approximates an ideal sensitivity function S_{id} with the spline-formed shape in bold in Figure C.8. The shape is originally given as a positive function W on the unit circle, and a normalizing factor $\rho > 0$ needs to be chosen so that $|S_{\text{id}}(e^{i\theta})| = \rho W(e^{i\theta})$ for $\theta \in [-\pi, \pi]$. An outer function h having the prescribed shape is given by

$$h(z) = \exp\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{e^{i\theta}+z}{e^{i\theta}-z} \log W(e^{-i\theta}) d\theta\right]$$

(see, e.g., [11, p. 62]). Now, define the function

$$f(z) = \rho \frac{z - \frac{1}{2}}{1 - \frac{1}{2}z} h(z),$$

where ρ is selected so that $f(0) = 1$. Then f is analytic in $\mathbb{D}$ and satisfies the interpolation conditions (C.23), and $S_{\text{id}}(z) = f(z^{-1})$. Clearly, f is nonrational and S_{id} represents a infinite-dimensional system.[1]

[1] For a systematic procedure to determine S_{id} from W for general interpolation constraints, see [12].

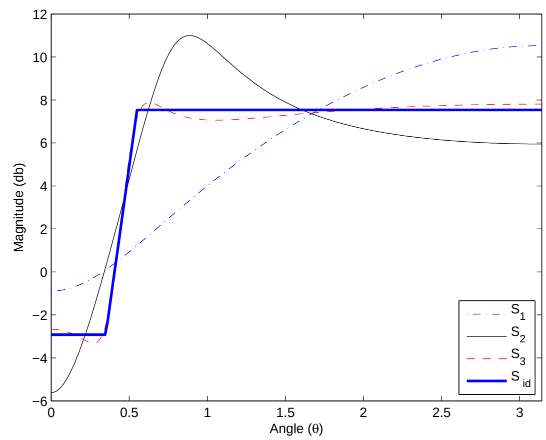

Figure C.8: Approximations of degree $1, 2$, and 3

By using the computational procedure in the beginning of the section, we determine the approximants g_n of f of degrees $n = 1, 2$ and 3 which satisfy the interpolation conditions (C.23). More precisely, g_1 is determined via steps (i) and (ii), whereas, for g_2 and g_3, we need to add one or two extra interpolation points and use (i) and (ii)′. That is, $z_0 = 0$ and $z_1 = 1/2$, and z_2 and z_2, z_3, respectively, are determined as in Remark C.3 with $w := f^{-1}$.

The magnitudes of the corresponding sensitivity functions S_1, S_2 and S_3, obtained from $S_n(z) = g_n(z^{-1})$, are depicted in Figure C.8. The degree of the controller corresponding to the approximant S_3 is two.

Example C.4. In [7] the problem of shaping the sensitivity function of a flexible beam with transfer function

$$P(s) = \frac{-6.4750s^2 + 4.0302s + 175.770}{s(5s^3 + 3.5682 + 139.5021s + 0.09290)}$$

is considered, and a controller is sought so that the sensitivity function is close to

$$S_{id} = \frac{s(s+1.2)}{s^2 + 1.2s + 1},$$

whose Bode plot is depicted in Figure C.9. The plant P has an unstable zero in 5.5308, a pole at 0 and has relative degree 2. For the controller to be strictly proper and the closed loop system to be internally stable, the interpolation condition

$$\begin{aligned} S(5.5308) &= S(\infty) = 1, \\ S(0) &= S(\infty)' = S(\infty)'' = 0, \end{aligned}$$

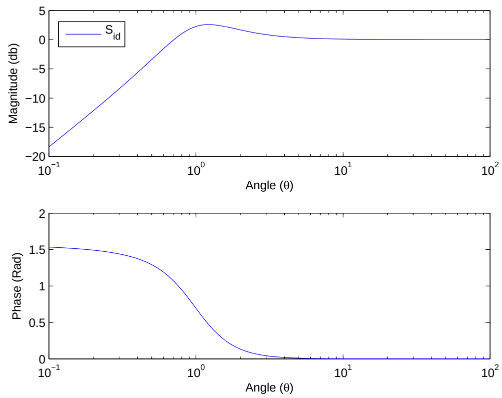

Figure C.9: *Bode plot of S_{id} and S*

needs to be satisfied.

In order to apply our theory as presented in this paper, we first transform the domain of the problem from $\mathbb{C}_+$ to $\mathbb{D}$, using the bilinear transformation

$$s \to z = \frac{s_0 - s}{s_0 + s}, \text{ where } s_0 = 3.1.$$

The constant $s_0 = 3.1$ is chosen the corresponding bilinear transformation maps the area of interest, $0.1i$ to $100i$, onto a large part of the unit circle. Choosing s_0 too small or too large might cause numerical problems. This yields

$$f_{id}(z) = S_{id}\left(s_0 \frac{1-z}{1+z}\right),$$

and the problem is then to find a stable function g that is close to $f_{id}(z)$ and which satisfies the constraints

$$\begin{aligned} g(-0.2816) &= g(-1) = 1, \\ g(1) &= g(-1)' = g(-1)'' = 0. \end{aligned} \quad \text{(C.24)}$$

However, f_{id} does not satisfy the constraints (C.24), and therefore the method in Section C.5 does not directly apply. Instead we would like to find an approximation f of f_{id} which satisfies the interpolation constraints, and then apply the degree reduction method on f.

Note that it is impossible to obtain an analytic function f which simultaneously satisfies the interpolation condition (C.24) and the criterion $|f(z)| \leq |f_{id}(z)|$ for $z \in \mathbb{T}$. If such a function f did exist, then $B := f/f_{id}$ would be analytic in $\mathbb{T}$ and bounded by one on $\mathbb{T}$. However,

$$B(-0.2816) = f(-0.2816)/f_{id}(-0.2816) = 1.0269 > 1,$$

and hence B violates the maximum principle. Therefore we need to be content with a function f which satisfies $|f(z)| \leq |f_{id}(z)|(1+\epsilon)$ for $z \in \mathbb{T}$ with some $\epsilon > 0.0269$.

If all the interpolation points of f were in $\mathbb{D}$, a straightforward method would be to take f as the minimizer of

$$\left\| \frac{f(z)}{f_{id}(z)} \right\|_\infty \text{ subject to (C.24).}$$

Then we would have $f = f_{id}B\alpha$, where B is a Blaschke product and $\alpha > 0$. But, since there are interpolation points on the boundary, a slightly larger region of analyticity need to be considered.

Note that f_{id} is analytic in $(1+\delta)\mathbb{D} := \{(1+\delta)z : |z| < 1\}$ for $0 < \delta < 0.44$, and let f be the function that minimizes

$$\left\| \frac{f(z)}{f_{id}(z)} \right\|_{H_\infty((1+\delta)\mathbb{D})}$$

subject to f satisfying the constraints (C.24). Now, for any $\epsilon > 0.0269$ one can find a $\delta > 0$ so that $\left\| \frac{f(z)}{f_{id}(z)} \right\|_\infty \leq 1+\epsilon$. We choose $\epsilon = 0.05$, and for this ϵ, $\delta = 0.05$ works.

Then the function f satisfies $|f(z)| \leq 1.05|f_{id}(z)|$ for $z \in \mathbb{T}$, and, since (C.24) holds for f it is possible to follow the steps (i) and (ii) to reduce the degree of f to 4. That is, let $(\rho, \sigma) \in \mathcal{K}_0 \times W_f$ be the minimizer of

$$\left\| 1 - \left| \frac{\rho}{\sigma} \right|^2 \right\|_\infty,$$

and let g be the unique function satisfying (C.24) and $\rho g \in \mathcal{K}$. Finally we transform the domain back to the continuous-time setting via

$$z \to s = s_0 \frac{1-z}{1+z},$$

whcih gives $S(s) = g\left(\frac{s_0 - s}{s_0 + s}\right)$ as depicted in Figure C.10.

Note that since there are interpolation points on the boundary, the relative H_2 bound is not meaningful. In fact, σ has poles in -1 that are not cancelled by zeros of f, and hence the right hand side of

$$\|\sigma(f-g)\|^2 \leq \frac{4\epsilon}{1-\epsilon}\|\sigma f\|^2.$$

will be infinite, rendering the inequality trivial. How to deal with interpolation points on the boundary in a more rigorous way will be the subject of further research.

Remark C.5. It is worth noting that if the main concern is a low order controller, one can consider a larger class of sensitivity functions with a possibility of better design. For clairity of presentation we will consider a discrete-time plant P. Briefly, we recall from [16] that

$$\deg C \leq \deg P + \deg S - n_p - n_z$$

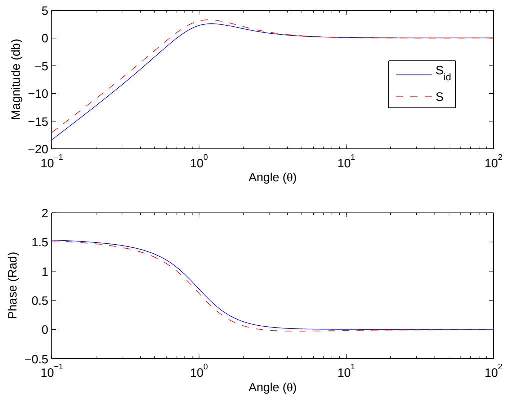

Figure C.10: *Bode plot of S_{id} and S*

where n_p and n_z are the number of unstable zeros and poles respectively of the plant P. Since the theory guarantees that $\deg S \leq n_p + n_z - 1$, the degree of the controller is less than the degree of the plant P. We then factor the transfer function of the plant into a stable and an unstable part as $(\beta_u\beta_s)/(\alpha_u\alpha_s)$, where β_u and α_u have roots in $\mathbb{D}$, and β_s and α_s have roots in $\mathbb{D}^C$. The idea is to use our knowledge about the stable part of the plant to construct a larger class of sensitivity functions for which the controller order is the same. Let

$$\mathcal{K}_{\alpha_s} = \left\{\sigma = \frac{b}{\tau\alpha_s}, b \in \text{Pol}(n + \deg\alpha_s), \sigma \text{ outer}\right\},$$

where

$$\tau(z) = \prod_{k=1}^{n_z}(1 - \bar{z}_k z)\prod_{k=1}^{n_p}(1 - \bar{z}_p z).$$

Now for any $\sigma \in \mathcal{K}_{\alpha_s}$ the minimizer of

$$\min \|\sigma S\| \text{ subject to } \begin{cases} S(z_k) = 1, & k = 0, \ldots, n_z \\ S(p_k) = 0, & k = 0, \ldots, n_p, \end{cases}$$

is of the form $S = \frac{\alpha_s a}{b}$, where $a \in \text{Pol}(n)$. Due to the interpolation constraints we have $\alpha_u | a$ and $\beta_u | (\alpha_s a - b)$, and hence

$$C = \frac{S-1}{PS} = \frac{\alpha_u(\alpha_s a - b)}{\beta_u\beta_s a} = \frac{\frac{\alpha_s a - b}{\beta_u}}{\frac{\beta_s a}{\alpha_u}},$$

where $\frac{\alpha_s a - b}{\beta_u}$ and $\frac{\beta_s a}{\alpha_u}$ are polynomials. Moreover, since $\deg \alpha_u + \deg \beta_u = n$, we have

$$\deg \frac{\alpha_s a - b}{\beta_u} \le n + \deg \alpha_s - \deg \beta_u = \deg \alpha_s + \deg \alpha_u$$
$$\deg \frac{\beta_s a}{\alpha_u} \le n + \deg \beta_s - \deg \alpha_u = \deg \beta_s + \deg \beta_u.$$

This shows that any choice of σ in the class $\mathcal{K}_{\alpha_s}$ will produce a controller of a degree less than the degree of the plant. By utilizing the stable part of the plant, we have shown that chosing sensitivity functions from a larger class will not increase the degree of the controller.

C.8 Concluding remarks

This paper presented a new theory for stability-preserving model reduction (for plants that need not be minimum-phase) that can also handle prespecified interpolation conditions and comes with error bounds. We have presented a systematic optimization procedure for choosing appropriate weight (and, if desired, interpolation points) so that the minimizer of a corresponding weighted H_2 minimization problem both matches the original system and has low degree.

The study of the H_2 minimization problem is motivated by the relation between the H_2 norm and the entropy functional used in bounded interpolation. Therefore, new concepts derived in this framework are useful for understanding entropy minimization. In fact, the degree reduction methods proposed in this paper easily generalize to the bounded case; see [12] for the method which preserves interpolation conditions. We are currently working on similar bounds for the positive real case; also, see [8].

C.9 Appendix

A quasi-convex optimization problem is an optimization problem for which each sublevel set is convex. The optimization problem to minimize (C.19), where p and q are polynomials of fixed degree is quasi-convex. For simplicity, we assume that f is real and hence that p and q are real as well.

As a first step, consider the *feasibility problem* of finding a pair (p, q) of polynomials satisfying

$$\left\| 1 - \left| \frac{qf}{p} \right|^2 \right\|_\infty \le \epsilon \tag{C.25}$$

for a given ϵ, or, equivalently,

$$-\epsilon |p(e^{i\theta})|^2 \le |p(e^{i\theta})|^2 - |q(e^{i\theta}) f(e^{i\theta})|^2 \le \epsilon |p(e^{i\theta})|^2$$

for all $\theta \in [-\pi, \pi]$. Since $|p|^2$ and $|q|^2$ are pseudo-polynomials, they have representations

$$\begin{aligned} |p(e^{i\theta})|^2 &= 1 + \sum_{k=1}^{n_p} p_k \cos(k\theta), \\ |q(e^{i\theta})|^2 &= \sum_{k=0}^{n_q} q_k \cos(k\theta), \end{aligned}$$

where n_p and n_q are the degree bounds on p and q respectively, and the first coefficient in $|p|^2$ is chosen to be one without loss of generality. Hence (C.25) is equivalent to

$$\begin{aligned} -1 - \epsilon \leq (1+\epsilon) \sum_{k=1}^{n_p} p_k \cos k\theta - |f(e^{i\theta})|^2 \sum_{k=0}^{n_q} q_k \cos k\theta, \\ 1 - \epsilon \leq (\epsilon - 1) \sum_{k=1}^{n_p} p_k \cos k\theta + |f(e^{i\theta})|^2 \sum_{k=0}^{n_q} q_k \cos k\theta, \end{aligned}$$

for all $\theta \in [-\pi, \pi]$. There is also a requirement on $1+\sum_{k=1}^{n_p} p_k \cos(k\theta)$ and $\sum_{k=0}^{n_q} q_k \cos(k\theta)$ to be positive. However, if $\epsilon \in (0,1)$, then the above constraints will imply positivity. The set of $p_1, p_2, \dots, p_{n_p}, q_0, q_1, \dots, q_{n_q}$ satisfying this infinite number of linear constraints is convex.

The most straightforward way to solve this feasibility problem is to relax the infinite number of constraints to a finite grid, which is dense enough to yield an appropriate solution. Here one must be carefully to check the positivity of $1 + \sum_{k=1}^{n_p} p_k \cos(k\theta)$ and $\sum_{k=0}^{n_q} q_k \cos(k\theta)$ in the regions between the grid points. Another method is the Ellipsoid Algorithm, described in detail in [2].

Minimizing (C.19) then amounts to finding the smallest ϵ for which the feasibility problem has a solution. This can be done by the the bisection algorithm, as described in [2]. Note that for $\epsilon = 1$, the trivial solution $q = 0$ is always feasible.

Acknowledgment

The authors would like to thank Tryphon Georgiou for many interesting discussions. Some of the ideas that lead to this paper originated in the join work [12] with him. The authors would also like to thank Martine Olivi for interesting discussions on how these methods compare with the gradient methods in [1, 14].

C.10 Bibliography

[1] L. Baratchart, M. Olivi, and F. Wielonsky. On a rational approximation problem in the real Hardy space H_2. *Theoretical Computer Science*, 94(2):175–197, 1992.

[2] S. Boyd, L. E. Ghaoui, E. Feron, and V. Balakrishnan. *Linear Matrix Inequalities in System and Control Theory*. SIAM Studies in Applied Mathematics, 1994.

[3] A. Bultheel and B. De Moor. Rational approximation in linear systems and control. *Journal of Computational and Applied Mathematics*, 121:355–378, 2000.

[4] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[5] C. I. Byrnes, T.T. Georgiou, A. Lindquist, and A. Megretski. Generalized interpolation in H-infinity with a complexity constraint. *Transactions of the American Mathematical Society*, 358:965–987, 2006.

[6] C. I. Byrnes and A. Lindquist. On the duality between filtering and Nevanlinna-Pick interpolation. *SIAM J. Control and Optimization*, 39:757–775, 2000.

[7] J. C. Doyle, B. A. Francis, and A. R. Tannenbaum. *Feedback Control Theory*. Macmillan Publishing Company, New York, 1992.

[8] G. Fanizza, J. Karlsson, A. Lindquist, and R. Nagamune. Passivity-preserving model reduction by analytic interpolation. *Linear Algebra and its Applications*, 425:608–633, 2007.

[9] A Gombani and G Michaletzky. On the parametrization of schur functions of degree n with fixed interpolating conditions. In *IEEE Proc. of CDC*, volume 4, pages 3875–3876, 2002.

[10] M. Green. A relative error bound for balanced stochastic truncation. *IEEE Trans. Automatic Control*, 33(10):961–965, 1988.

[11] K. Hoffman. *Banach Spaces of Analytic Functions*. Dover Publications, 1962.

[12] J. Karlsson, T. Georgiou, and A Lindquist. The inverse problem of analytic interpolation with degree constraint. In *IEEE Proc. CDC*, San Diego, 2006.

[13] D. G. Luenberger. *Optimization by Vector Space Methods*. John Wiley & Sons, 1969.

[14] J. Marmorata and M. Olivi. Nudelman interpolation, parametrizations of lossless functions and balanced realizations. *Automatica*, 43(8):1329–1338, 2007.

[15] D. Mustafa and K. Glover. *Minimum Entropy H-infinity Control*. Springer-Verlag, Berlin Heidelberg, 1990.

[16] R. Nagamune. *Robust Control with Complexity Constraint: A Nevanlinna-Pick Interpolation Approach*. PhD thesis, Optimization and System Theory, Department of Mathematics, Royal Institute of Technology (KTH), Stockholm, Sweden, 2002.

[17] J. R. Partington. Some frequency-domain approaches to the model reduction of delay systems. *Annual Reviews in Control*, 28:65–73, 2004.

[18] M.G. Safonov and R.Y.Chiang. Model reduction for robust control: A schur relative error method. *International J. of Adaptive Control and Signal Processing*, 2:259–272, 1988.

[19] S. Skogestad and I. Postlethwaite. *Multivariate Feedback Control*. John Wiley & Sons, 1996.

[20] K. C. Sou, A. Megretski, and L. Daniel. A quasi-convex optimization approach to parameterized model order reduction. *Proc. IEEE Design Automation Conference*, pages 933–938, 2005.

[21] M. S. Takyar, A. Nasiri Amini, and T. T. Georgiou. Weight selection in interpolation with a dimensionality constraint. In *Proc. IEEE CDC*, pages 3536–3541, 2006.

[22] W. Wang and M. G. Safonov. Relative-error bound for discrete balanced stochastic truation. *International Journal of Control*, 54(3):593–612, 1991.

[23] K. Zhou. Frequency-weighted l_∞ norm and optimal Hankel norm model reduction. *IEEE Trans. Automat. Control*, 40:1687–1699, 1995.

[24] K. Zhou and J.C. Doyle. *Essentials of Robust Control*. Prentice-Hall, Inc, 1998.

Localization of Power Spectra

Johan Karlsson and Tryphon Georgiou

Abstract

The purpose of this paper is to study the topology and develop metrics that allows for localization of power spectra, based on second-order statistics. We show that the appropriate topology is the weak*-topology and give several examples on how to constuct such metrics. This allows us to quantify uncertainty of spectra in an natural way and to calculate a priori bounds on spectral uncertainty, based on second-order statistics. Finally, we study identification of spectral densities and relate this to the tradeoff between resolution and variance of estimates.

D.1 Introduction

The estimation of power spectra of stationary time-series relies on second order statistics. The premise is that these are moments of an underlying power spectral distribution —the power spectrum. Thus, the question arises as to how much is "knowable" about the power spectrum from such statistics. In other words, in what ways statistics localize the power spectrum. Traditionally, there have been "methods" that lead to specific power spectra which, in one way or another, are consistent with the recorded data and estimated moments. For instance, the correlogram, the periodogram, Burg's algorithm, and the maximum entropy spectrum are specific such choices [12, 20]. In general, there exists a large family of admissible power spectra. Bounding the "values" of the spectral density function at a specific region based on knowledge of a finite set of statistics is an ill-posed problem. The proper notion of localization is a certain weak notion of bounding the energy over parts of the frequency band. Thus, the goal of this paper is to study the appropriate topology and develop suitable metrics. These metrics allow localization of power spectra and they can be used to quantify spectral uncertainty on the basis of estimated statistics.

A typical set of statistics for a stationary stochastic process consists of finitely many covariance samples. Throughout, we consider such a process $\{y_t \,:\, t \in \mathbb{Z}\}$ to be a discrete-time, zero-mean, second-order stationary stochastic process. The covariances (equivalently, autocorrelation samples)

$$c_k := \mathcal{E}\{y_t \bar{y}_{t-k}\}, \text{ for } k = 0, \pm 1, \pm 2, \ldots, \pm n,$$

where $\mathcal{E}\{\cdot\}$ denotes the expectation operator, provide moment constraints for the power spectrum $d\mu$ of the process:

$$c_k = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-ik\theta}d\mu(\theta) \text{ for } k = 0, \pm 1, \pm 2 \ldots, \pm n. \tag{D.1}$$

The power spectrum is a non-negative measure on the unit circle $\mathbb{T} = \{z = e^{i\theta} : \theta \in (-\pi, \pi]\}$ which, for simplicity, we identify with the interval $(-\pi, \pi]$. We use the symbol $\mathfrak{M}$ to denote the class of such measures. The problem of determining $d\mu$ from the covariance samples (finitely or infinitely many) is known as the trigonometric moment problem. Classical theory on this problem originates in the work of Toeplitz and Carathéodory at the turn of the 20th century and has evolved into a rather deep chapter of functional analysis and of operator theory [1, 18, 10, 5, 2]. The classical monograph by Geronimus [10] presents a wide range of results on the asymptotic convergence of solutions to the trigonometric moment problem, which include the asymptotic behavior of the maximum entropy spectrum [10, Theorem 5.7] and of spectral envelops [10, Theorem 5.7] (c.f. [4, 12, 8]). It also studies at length the convergence of the corresponding spectral factors.

In the present work we attempt to address certain question of practical interest. Invariably, the covariance samples are estimated from a finite observation record of the process, and are known with limited accuracy. Thus, in a typical experiment, as the observation record increases so does the accuracy and the length of the covariance estimates. In this context, we seek quantitative metrics of spectral uncertainty that have the following properties:

1. given a finite set of covariance samples, the family of consistent power spectra has a finite diameter,

2. the diameter of the uncertain set of power spectra shrinks to zero as the accuracy of the covariance samples increases and as their number tends to infinity.

The latter condition is dictated by the fact that the trigonometric moment problem is known to be determined, i.e., there is a unique power spectrum consistent with an infinite sequence of covariances. As we will explain, the proper topology which allows for these properties to hold is the weak* topology on measures (cf., [13, page 8]). There is a variety of metrics that can be used to metrize the topology, and thus, in principle, quantify spectral uncertainty. A main contribution of this work is to present a class of metrics for which the radius of spectral uncertainty is computable given a finite set of statistics.

In Section D.2 we review the trigonometric moment problem and discuss relevant relations with complex analysis and functional analysis. In Section D.3 derive the connection between uncertainty sets and weak* continuous metrics and state the main theorem. In Section D.4 give several examples of weak*-continuous metrics. In Section D.5 the we calculate the size of the uncertainty for a metric and show that it satisfies the desired limit properties. In Section D.6 we study an identification example and in Section D.7 discuss conclusions and further directions.

D.2 On the trigonometric moment problem

The covariances c_k, $k = 0, \pm 1, \pm 2, \ldots$, of the random process $\{y_t \,:\, t \in \mathbb{Z}\}$ are the Fourier coefficients of the spectral measure $d\mu$ as in (D.1). These are characterized by the non-negativity of the Toeplitz matrices [11, 12]

$$T_n = \begin{bmatrix} c_0 & c_{-1} & \cdots & c_{-n} \\ c_1 & c_0 & \cdots & c_{-n+1} \\ \vdots & \vdots & \ddots & \vdots \\ c_n & c_{n-1} & \cdots & c_0 \end{bmatrix},$$

for $n = 0, 1, \ldots$. When $T_n \geq 0$ and singular for a particular value n, then $\mathrm{rank}(T_{n+\ell}) = \mathrm{rank}(T_n)$ for all $\ell \geq 1$. In this case, $d\mu$ is singular with respect to the Lebesgue measure and consists of finitely many "spectral lines," equal in number to $\mathrm{rank}(T_n)$ [11, page 148]. Because $d\mu$ is a real measure, $c_k = \bar{c}_{-k}$ for $k = 0, 1, \ldots$, hence we use only positive indices and we will refer by

$$\mathbf{c}_{0:n} := (c_0, c_1, \ldots, c_n)$$

to the vector of the first $(n+1)$ moments, and by

$$\mathbf{c} := (c_0, c_1, \ldots)$$

to the infinite sequence. The sequence $\mathbf{c}$ is said to be *positive* if $T_n > 0$ for all n. Similarly $\mathbf{c}_{0:n}$ is said to be *positive* if $T_n > 0$. Accordingly, the term *non-negative* is used when the relevant Toeplitz matrices are non-negative definite.

As noted in the introduction, the power spectrum of a discrete-time stationary process is a bounded non-negative measure on the unit circle. The derivative (of its absolutely continuous part) is referred to as the spectral density function, while the singular part typically contains jumps (spectral lines) associated with the presence of sinusoidal components. In general, the singular part may have a more complicated mathematical structure that allocates "energy" on a set of measure zero without the need for distinct spectral lines [11, page 5]. From a mathematical viewpoint such spectra are important as they represent limits of more palatable spectra, and hence, represent a form of completion. Non-negative measures are naturally associated with analytic and harmonic functions—a connection which has profitably been exploited in classical circuit theory in the context of passivity. More specifically, power spectra are, in a very precise sense, boundary limits of the (harmonic) real parts of so-called "positive-real functions." Below we summarize relevant concepts.

Herglotz' theorem [1] states that if $d\mu$ is a bounded non-negative measure on $\mathbb{T}$, then

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

is analytic in $\mathbb{D} := \{z \,:\, |z| < 1\}$ and the real part is non-negative. Such functions are referred to as either "positive-real" or, as Carathèodory functions. Conversely, any positive-real function can be represented (modulo an imaginary constant) by the above formula for a suitable non-negative measure. The Poisson integral of a non-negative measure $d\mu$

$$P[d\mu](z) := \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(t-\theta)d\mu(\theta), \quad z = re^{it},$$

where $P_r(\theta) = \frac{1-r^2}{|1-re^{i\theta}|^2}$ is the Poisson kernel, is a harmonic function which is non-negative in $\mathbb{D}$ and is equal to the real part of $H[d\mu](z)$.

We denote by $C(\mathbb{T})$ the class of real valued continuous functions on $\mathbb{T}$. It is quite standard that the dual space of $C(\mathbb{T})$, i.e the space of bounded linear functionals $\Lambda : C(\mathbb{T}) \to \mathbb{R}$, can be identified with the space of bounded measures on $\mathbb{T}$ ([13, page 7], [16, page 374]. This is the Riesz representation theorem, which asserts the existence of bounded measure $d\mu$ such that

$$\Lambda(f) = \int_{\mathbb{T}} f(t)d\mu(t)$$

for all $f \in C(\mathbb{T})$. Thus, for any two measures that are different, there exists a continuous function that the two measures integrate to different values. In other words, continuous functions serve as "test functions" to differentiate between measures, and bounds on such integrals define the weak* topology. A sequence of measures $d\mu_n$, $n = 1, 2, \ldots$, converges to $d\mu$ in the weak* topology if $\int f d\mu_n \to \int f d\mu$ for every $f \in C(\mathbb{T})$. The limit, when it exists, is defined uniquely by the sequence.

Given either a positive-real function $H(z)$, or its real part $P(z)$, the measure $d\mu$ such that $H(z) = H[d\mu](z)$ and $P(z) = P[d\mu](z)$ is uniquely determined by the limit of $P(re^{i\theta})d\theta \to d\mu$ as $r \to 1$ in the weak* topology [13, page 33]. Similarly, given a positive sequence **c**, the measure $d\mu$ can be determined as the limit in the weak* topology of finite Fourier sums or Cesaro means [13, page 24], to which we will return in Section D.6.

D.3 Localization of power spectra

We postulate a situation where covariances $\mathbf{c}_{0:n}$ of the stochastic process $\{y_k\}_{k\in\mathbb{Z}}$, which has power spectrum $d\mu$, are estimated with an absolute error bounded by ϵ_n. Typical assumptions on the nature of power spectra such as, autoregressive or smooth, may in general not be justified. In the absence of such information, $d\mu$ may be any power spectrum in the uncertainty set

$$\mathcal{F}_{\mathbf{c}_{0:n},\epsilon_n} = \left\{ d\mu \geq 0 : \left| c_k - \int_{-\pi}^{\pi} e^{-ik\theta} d\mu \right| < \epsilon_n, k = 0, 1, \ldots, n \right\}.$$

In the limit, as the number of covariances increases and as the measurement errors decrease to zero, the uncertainty set shrinks to a unique power spectrum. This is due to the fact that an infinite limit sequence **c** defines a unique power spectrum, since the trigonometric problem is determined.

Thus, we seek suitable metrics δ on the space of positive measures $\mathfrak{M}$ that provide a meaningful and computationally tractable notion of "diameter"

$$\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n},\epsilon_n}) := \sup\{\delta(d\mu_0, d\mu_1) \ : \ d\mu_0, d\mu_1 \in \mathcal{F}_{\mathbf{c}_{0:n},\epsilon_n}\},$$

so as to quantify "modeling uncertainty" and determine the size of $\mathcal{F}_{\mathbf{c}_{0:n},\epsilon}$ in the appropriate topology. The diameter should reflect shrinkage to a singelton via

$$\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n},\epsilon_n}) \to 0 \text{ as } n \to \infty, \epsilon_n \to 0. \tag{D.2}$$

This condition forces the underlying metric to be weak* continuous, as we will se next.

Theorem D.1. *The metric δ satisfies* (D.2) *if and only if it δ is weak* continuous.*

Proof. See the Appendix. □

We now consider the case where the finite covariance sample $\mathbf{c}_{0:n}$ is known. If $\mathbf{c}_{0:n}$ is positive, then the uncertainty set

$$\mathcal{F}_{\mathbf{c}_{0:n}} := \{d\mu : \ d\mu \geq 0, \text{ and (D.1) holds}\}$$

contains infinitely many power spectra. If $\mathbf{c}_{0:n}$ is only non-negative, and hence T_n is singular, then the family $\mathcal{F}_{\mathbf{c}_{0:n}}$ consists of a single power spectrum $d\mu$ [11, page 148]. The following two results are corollaries of Theorem D.1.

Corollary D.2. *Let $\mathbf{c}$ be a non-negative sequence and let $\delta(\cdot,\cdot)$ be a weak*-continuous metric. Then, as $n \to \infty$, $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}}) \to 0$.*

Proof. The corollary follows by virtue of the fact that

$$\mathcal{F}_{\mathbf{c}_{0:n}} \subset \mathcal{F}_{\mathbf{c}_{0:n},\epsilon_n},$$

or independently from [10, §1.16] in view of Proposition D.5. □

Corollary D.3. *Let $\mathbf{c}_{0:n}$ be a non-negative sequence such that T_n is a singular Toeplitz matrix, and let $\delta(\cdot,\cdot)$ be a weak*-continuous metric. If $\hat{\mathbf{c}}_{0:n}(k)$ ($k = 1, 2, \ldots$) are non-negative $(n+1)$-sequences of moments tending to $\mathbf{c}_{0:n}$, then $\rho_\delta(\mathcal{F}_{\hat{\mathbf{c}}_{0:n}(k)}) \to 0$ as $k \to \infty$.*

For an independent proof, see [15].

Remark D.1. The total variation ($\int |d\mu_0 - d\mu_1|$) is not weak* and therefore the conclutions of the two corollaries would fail if this was used as the metric. To see this, note that if $\mathbf{c}_{0:n}$ is positive, then $\mathcal{F}_{\mathbf{c}_{0:n}}$ contains infinitely many measures and among them at least two singular measures with non-overlapping support, i.e., $\mathrm{s}upp(d\mu_0) \cap \mathrm{s}upp(d\mu_1) = \emptyset$ [18]. The total variation of their difference is $2c_0$. □

D.4 Weak*-continuous metrics

In general, a finite set of second order statistics cannot dictate the precise value of the power spectrum locally. Indeed, given any finite positive sequence $\mathbf{c}_{0:n}$ and any $\theta_0 \in (-\pi, \pi]$, then for any value $\alpha \geq 0$ there exists an $\epsilon > 0$ and an absolutely continuous measure $d\mu = f d\theta \in \mathcal{F}_{\mathbf{c}_{0:n}}$ such that

$$f(\theta) = \alpha \text{ for } \theta \in (\theta_0 - \epsilon, \theta_0 + \epsilon).$$

What can be said instead, is that the range of values

$$\left\{ \int_{\mathbb{T}} g d\mu \; : \; d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}} \right\}, \tag{D.3}$$

for any particular test function $g \in C(\mathbb{T})$, is bounded. Furthermore, as $n \to \infty$, this range tends to zero. In fact, due to weak*-continuity, the range of values tend to zero for any of the scenarios in Theorem D.1 and its two corollaries. Finding the maximum and the minimum of (D.3) is a linear programming problem on an infinite dimensional domain. Provided g is symmetric around 0 and the covariance sequence $\mathbf{c}_{0:n}$ is real, the dual problems, which give the lower and upper bounds of (D.3), are

$$\begin{aligned} &\max \left\{ \mathbf{c}_{0:n}\lambda \; : \; \textstyle\sum_{k=0}^{n} \lambda_k \cos(k\theta) \le g(\theta), \theta \in (-\pi, \pi] \right\}, \\ &\min \left\{ \mathbf{c}_{0:n}\lambda \; : \; g(\theta) \le \textstyle\sum_{k=0}^{n} \lambda_k \cos(k\theta), \theta \in (-\pi, \pi] \right\}, \end{aligned}$$

where $\lambda^T = (\lambda_0,\, \lambda_1,\, \ldots,\, \lambda_n)$.

A class of weak*-continuous metrics can be sought in the form

$$\delta(d\mu_0, d\mu_1) = \sup_{\xi \in K} \left| \int_{\mathbb{T}} g_\xi (d\mu_0 - d\mu_1) \right|, \tag{D.4}$$

for $\{g_\xi\}_{\xi\in K} \subset C(\mathbb{T})$, provided the family $\{g_\xi\}_{\xi\in K}$ is sufficiently rich to distinguish between measures and yet, small enough so that continuity is ensured. The precise conditions are given next.

Proposition D.4. *The functional $\delta(d\mu_0, d\mu_1)$ defined in* (D.4) *is a weak*-continuous metric if and only if the following two conditions hold:*

1. *for any two measures $d\mu_0, d\mu_1 \in \mathfrak{M}$, there is a $\xi \in K$ such that $\int_{\mathbb{T}} g_\xi d\mu_0 \neq \int_{\mathbb{T}} g_z d\mu_1$, and*

2. *the set $\{g_\xi\}_{\xi\in K}$ is relatively compact in $C(\mathbb{T})$.*

Proof. It is clear that condition 1 holds if and only if $\delta(d\mu_0, d\mu_1)$ is positive whenever $d\mu_0 \neq d\mu_1$. The triangle inequality and symmetry always holds for such δ, so we only need to show that condition 2 holds if and only if δ is weak*-continuous.

We will show that condition 2 implies that δ is weak*-continuous by contradiction. Assume therefore that condition 2 holds, but that δ is not weak*. Then there exists $d\mu_k \to d\mu$ in weak* such that $\delta(d\mu_k, d\mu) > \epsilon$, $k = 1, 2, \ldots$, and hence there exists $g_{\xi_k}, \xi_k \in K$, such that

$$\epsilon < \left| \int_{\mathbb{T}} g_{\xi_k} (d\mu_k - d\mu) \right|, \; k = 1, 2, \ldots.$$

Since the set $\{g_\xi\}_{\xi\in K}$ is relatively compact in $C(\mathbb{T})$, there is a subsequence $(g_\ell, d\mu_\ell)$ of $(g_{\xi_k}, d\mu_k)$ such that $g_\ell \to g \in C(\mathbb{T})$. A contradiction follows, since

$$\begin{aligned}\epsilon &< \left|\int_{\mathbb{T}} g_\ell(d\mu_\ell - d\mu)\right| \\ &\leq \|g_\ell - g\|_\infty \int_{\mathbb{T}} |d\mu_\ell - d\mu| + \left|\int_{\mathbb{T}} g(d\mu_\ell - d\mu)\right| \\ &\to 0 \text{ as } \ell \to \infty,\end{aligned}$$

and hence δ is weak*-continuous whenever condition 2 holds.

A well known result of Arzelà (see e.g., [16, page 102]) is that a set of functions is relatively compact in $C(\mathbb{T})$ if and only if the set of functions is uniformly bounded and equicontinuous. If $\{g_\xi\}_{\xi\in K}$ is not equicontinuous, then there exists an $\epsilon > 0$ such that for any $k = 1, 2, \ldots$ one can find $\theta_k, \phi_k \in \mathbb{T}$, and $\xi_k \in K$, that satisfies

$$|\theta_k - \phi_k| < \frac{1}{k} \text{ and } |g_{\xi_k}(\theta_k) - g_{\xi_k}(\phi_k)| > \epsilon. \tag{D.5}$$

Let (θ_ℓ, ϕ_ℓ) be a subsequence of (θ_k, ϕ_k) such that $\theta_\ell \to \theta_0 \in \mathbb{T}$ as $\ell \to \infty$, and let $d\mu_\ell$ and $d\nu_\ell$ be the measurs that consist of a unit mass in θ_ℓ and ϕ_ℓ, respectively. From (D.5) it follows that $\phi_\ell \to \theta_0$, and hence that $d\mu_\ell \to d\mu_0$ and $d\nu_\ell \to d\mu_0$ in weak*, where $d\mu_0$ is the measure that consist of a unit mass in θ_0. From (D.5) it follows that

$$\begin{aligned}\delta(d\mu_\ell, d\mu_0) + \delta(d\nu_\ell, d\mu_0) &\geq \delta(d\mu_\ell, d\nu_\ell) \\ &\geq |g_{\xi_\ell}(\theta_\ell) - g_{\xi_\ell}(\phi_\ell)| > \epsilon.\end{aligned}$$

From this, it is evident that δ is not weak*-continuous since both $\delta(d\mu_\ell, d\mu_0)$ and $\delta(d\nu_\ell, d\mu_0)$ cannot converge to 0.

Similarly, if $\{g_\xi\}_{\xi\in K}$ is not uniformly bounded, then for any $k = 1, 2, \ldots$ one can find $\theta_k \in \mathbb{T}$ and $\xi_k \in K$ such that

$$|g_{\xi_k}(\theta_k)| > k. \tag{D.6}$$

Let $d\mu_k$ be the measures that consist of a unit mass in θ_k. the metric δ is not weak*-continuous since $\frac{1}{k}d\mu_k \to 0$ in weak*, but $\delta(\frac{1}{k}d\mu_k, 0) > 1$ for all k. □

Condition 1 ensures positivity and condition 2 ensures weak*-continuity. The triangle inequality and symmetry always holds for such δ.

D.4.1 Metrics based on smoothing

A simple way to devise weak*-continuous metrics on measures is by first smoothing the measures via convolution with a fixed suitable continuous function, and then compare the spectral density functions of the smoothed spectra. The function must have non-zero Fourier coefficients, otherwise it will not differentiate between certain measures. Thus, if $g \in C(\mathbb{T})$ is such a function, then

$$\delta_{\text{smooth},g}(d\mu_0, d\mu_1) = \|g * (d\mu_0 - d\mu_1)\|_\infty,$$

is a weak*-continuous metric. Here,

$$(g * d\mu)(\xi) = \int_{-\pi}^{\pi} g(\xi - \theta) d\mu(\theta),$$

denotes the circular convolution and $\| \cdot \|_\infty$ the L_∞ norm. It follows from Proposition D.4 that $\delta_{\text{smooth},g}(d\mu_0, d\mu_1)$ is a weak*-continuous metric. To see this, note that

$$\|g * (d\mu_0 - d\mu_1)\|_\infty = \sup_{\xi \in (-\pi,\pi]} \left| \int_{-\pi}^{\pi} g(\xi - \theta)(d\mu_0(\theta) - d\mu_1(\theta)) \right|,$$

and hence, condition $2)$ of the proposition holds. Then, if $g(\theta) = \sum_{k=-\infty}^{\infty} g_k e^{ik\theta}$ and if $(\ldots, a_{-1}, a_0, , a_1, \ldots)$ are the Fourier coefficients of $d\mu_0(\theta) - d\mu_1(\theta)$, then

$$g * (d\mu_0 - d\mu_1)(\xi) = \sum_{k=-\infty}^{\infty} g_{-k} a_k e^{ik\xi}.$$

Since $g_k \neq 0$ for all $k \in \mathbb{Z}$, the above expression cannot vanish identically unless all the a_k's are zero, in which case $d\mu_0 = d\mu_1$.

D.4.2 Metrics based on transportation

The Monge-Kantorovic transportation problem, can be put into this framework. The problem amounts to minimizing the cost of transportation between two distributions of equal mass, e.g., $d\mu_0$ and $d\mu_1$ where $\int_{\mathbb{T}} d\mu_0 = \int_{\mathbb{T}} d\mu_1$. In this, a transportation plan $d\pi(\theta, \phi)$ is sought which corresponds to a non-negative distribution on $\mathbb{T} \times \mathbb{T}$ and is such that

$$\int_{\theta \in \mathbb{T}} d\pi(\theta, \phi) = d\mu_0(\phi) \text{ and } \int_{\phi \in \mathbb{T}} d\pi(\theta, \phi) = d\mu_1(\theta). \tag{D.7}$$

Then, the minimal cost

$$\min \left\{ \int_{\mathbb{T}\times\mathbb{T}} |\theta - \phi| d\pi(\theta, \phi) \ : \ d\pi \text{ satisfies (D.7)} \right\}$$

is the Wasserstein-1 distance between $d\mu_0$ and $d\mu_1$, and is a weak*-continuous metric (see e.g., [22, chapter 7]). This problem admits a dual formulation, referred to as the Kantorovic duality (see [22]):

$$W_1(d\mu_0, d\mu_1) = \max_{\|g\|_L \leq 1} \int g(d\mu_0 - d\mu_1),$$

where $\|f\|_L = \sup_{\theta,\phi} \frac{|f(\theta) - f(\phi)|}{|\theta - \phi|}$ denotes the Lipschitz norm.

Power spectra, in general, cannot be expected to have the same total mass. In this case, $\delta_{1,\kappa}(d\mu_0, d\mu_1)$ defined by

$$\inf_{\int d\nu_0 = \int d\nu_1} W_1(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^{1} \int_{\mathbb{T}} |d\mu_i - d\nu_i|, \tag{D.8}$$

is a weak*-continuous metric for an arbitrary but fixed $\kappa > 0$. The interpretation is that $d\mu_0$ and $d\mu_1$ are perturbations of the two measures $d\nu_0$ and $d\nu_1$, respectively, with equal mass. Then, the cost of transporting $d\mu_0$ and $d\mu_1$ to one another can be thought of as the cost of transporting $d\nu_0$ and $d\nu_1$ to one another plus the size of the respective perturbations. This is introduced in [6] and the metric admits a dual formulation

$$\delta_{1,\kappa}(d\mu_0, d\mu_1) = \max_{\substack{\|g\|_\infty \leq \kappa \\ \|g\|_L \leq 1}} \int g(d\mu_0 - d\mu_1),$$

which is in the form of the Proposition D.4. Various other generalizations of transportation distance that apply to power spectra are also being proposed and studied in [6].

D.4.3 Metrics based on the Poisson kernel

In this section we define metrics based on the Poisson integral,

$$P[d\mu](z) = \frac{1}{2\pi}\int_{-\pi}^{\pi} P_r(t-\theta)d\mu(\theta), \quad z = re^{it},$$

which allow computation of explicit bounds on uncertainty sets (see Section D.5). It is easy to see that weak*-convergence of measures is equivalent to certain types of convergense of their harmonic counterpart, as stated next.

Proposition D.5. *Let $\{d\mu_k\}_{k=1}^{\infty}$ be a sequence of uniformly bounded signed measures on $\mathbb{T}$, let $d\mu$ be a bounded measure on $\mathbb{T}$, and let $u(z) = P[d\mu](z)$, $u_k(z) = P[d\mu_k](z)$ be their corresponding Poisson integrals. The following statements are equivalent:*

1. $d\mu_k \to d\mu$ *weak*,*
2. $u_k(z) \to u(z)$ *pointwise* $\forall z \in \mathbb{D}$,
3. $u_k(z) \to u(z)$ *in* $L_1(\mathbb{D})$,
4. $u_k(z) \to u(z)$ *uniformly on compact subsets of* $\mathbb{D}$.

Proof. See Appendix. □

The maximal distance between the harmonic functions, on a compact non-finite set $K \subset \mathbb{D}$, gives rise to a weak*-continuous metric

$$\delta_K(d\mu_0, d\mu_1) = \sup_{z\in K} |P(d\mu_0 - d\mu_1)(z)|. \tag{D.9}$$

This is clear, since the resulting family of the Poisson kernels satisfies the properties in Proposition D.4.

Remark D.2. In practice, it is often the case that one is interested in comparing spetra over selected frequency bands. To this end, various schemes have been consider which rely on pre-processing with a choice of "weighting" filters and filter banks (see e.g., [4], [21], and [3, 9]). The choice of the point-set K in (D.9) can be used to dictate the resolution of the metric over such frequency bands. To see how this can be done, consider K to designate an arc $\{\xi = re^{i\theta} \,:\, \theta \in [\theta_0 - \epsilon, \theta_0 + \epsilon]\}$. This satisfies the condtions of Proposition D.4 and thus, δ_K is a weak* metric. At the same time, the values $P(d\mu)(\xi)$, with $\xi \in K$, represent the variance at the output of a filter with transfer function $z/(z-\xi)$. These are bandpass filters with a center frequency $\arg(\xi)$ and bandwidth which depends on the choice of r. Thus, the metric compares the respective variance after the spectra have been weighed by a continium (for $\xi \in K$) of such frequency selective bank of filters. □

D.5 The size of the uncertainty set

The size of the uncertainty set with respect to the distance δ_K turns out to be especially easy to compute. Indeed, the diameter is attained on a special subset of the essential boundary which corresponds to measures with only $n+1$ points of increase (i.e., support). This is the content of the following proposition.

Proposition D.6. *Let* $\mathbf{c}_{0:n}$ *be a positive covariance sequence and let* $K \subset \mathbb{D}$ *be closed. Then*

$$\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}}) = \max_{z \in K} \left\{ 2 \left(\left| \frac{\frac{2}{1-z\bar{z}} + (b_z, d_z)}{(b_z, b_z)} \right|^2 - \frac{(d_z, d_z)}{(b_z, b_z)} \right)^{\frac{1}{2}} \right\},$$

where

$$b_z = \begin{pmatrix} z^{-1} \\ z^{-2} \\ \vdots \\ z^{-n-1} \end{pmatrix},\ d_z = \begin{pmatrix} z^{-1}(c_0) \\ z^{-2}(c_0 + 2c_1 z) \\ \vdots \\ z^{-n-1}(c_0 + 2c_1 z + \cdots + 2c_n z^n) \end{pmatrix},$$

and (x, y) *denote the inner product* $y^* T_n^{-1} x$. *Furthermore,* $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ *is attained as the distance between two elements of* $\mathcal{F}_{\mathbf{c}_{0:n}}$ *which are both singular with support containing at most* $n+1$ *points.*

Proof. See appendix. □

Both claims in Proposition D.6 can be used separately for computing $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$. The first one suggests finding the maximum of a real valued function over K. The second claim suggests search for a maximum for $\delta_K(d\mu_1, d\mu_2)$ over a rather small subset of e$xt(\mathcal{F}_{\mathbf{c}_{0:n}})$, namely nonnegative sequences $\mathbf{c}_{0:(n+1)}$ parametrized by c_{n+1} being a solution of the quadratic equation

$$\det(T_{n+1}) = 0.$$

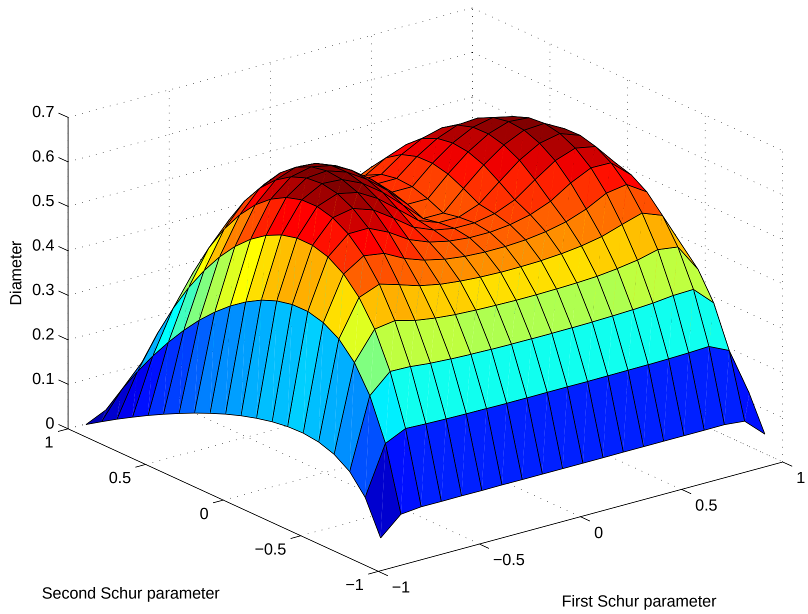

Figure D.1: ρ_{δ_K} *as a function of* γ_1, γ_2 *when* $c_0 = 1$. $K = \{z : |z| \leq 0.5\}$

The corresponding values for c_{n+1} lie on a circle in the complex plane, and hence, computation of $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ will require search on a torus (each of the two extremal $d\mu_1$, $d\mu_2$ where the diameter is attained can be thought of as points on the circle).

As an example, Figure D.1 shows $\rho_{\delta_K}(\mathcal{F}_{\mathbf{c}_{0:n}})$ for

$$\mathbf{c}_{0:2} = (1,\ c_1,\ c_2)$$

as a function of the corresponding Schur parameters [10]

$$\begin{aligned} -1 <& \quad \gamma_1 := c_1 & < 1, \\ -1 \leq& \quad \gamma_2 := \frac{\det \begin{pmatrix} c_1 & c_2 \\ 1 & c_1 \end{pmatrix}}{\det \begin{pmatrix} 1 & c_1 \\ \bar{c}_1 & 1 \end{pmatrix}} & \leq 1, \end{aligned}$$

and K is taken as $\{z\ :\ |z| \leq 0.5\} \subset \mathbb{D}$.

The plot confirms that the diameter decreases to zero as the parameters or, alternatively, the covariances c_1 and c_2, tend to the boundary of the "positive" region (which in the Schur coordinates corresponds to the unit square). However, it is interesting to note that the diameter of $\mathcal{F}_{\mathbf{c}_{0:n}}$ as a function of $\mathbf{c}_{0:n}$ has several local maxima.

Remark D.3. Computation of the diameter $\rho_\delta(\mathcal{F}_{\mathbf{c}_{0:n}})$ of the uncertainty set amounts to solving the infinite-dimensional optimization problem

$$\sup\{\delta(d\mu_1, d\mu_2) \,:\, d\mu_1, d\mu_2 \in \mathcal{F}_{\mathbf{c}_{0:n}}\}. \tag{D.10}$$

If δ is a weak*-continuous and jointly convex function, then the diameter is attained as the precise distance between two elements which are extreme points $\mathcal{F}_{\mathbf{c}_{0:n}}$. Extreme points are the points with the property that they are not a nontrivial convex combination of elements in the set, denoted e$xt(\cdot)$. Then, $d\mu \in$ e$xt(\mathcal{F}_{\mathbf{c}_{0:n}})$ if and only if $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$ and the support of $d\mu$ consists of at most $2n+1$ points (see [15]). Thus, e$xt(\mathcal{F}_{\mathbf{c}_{0:n}})$ admits a finite dimensional characterization and (D.10) reduces to a finite dimensional problem.

D.6 Identification in weak*

In this example we consider identification of a spectrum based on covariances. Although the uncertainty set contains a large set of spectra which may have considerably different qualitatively properties, they are all close in the weak*-topology. In this context, the trade-off between variance and resolution can be studied. The resolution is specified by the metric (in this case by the set K), and then sufficient data has to be provided to ensure that the size of the uncertainty set is small.

Consider the stochastic process

$$y_t = \cos(0.5t + \varphi_1) + \cos(t + \varphi_2) + w_t + \frac{1}{3} w_{t-1}$$

where w_t is a white noise process and φ_1, φ_2 are random variables with uniform distribution on $(-\pi, \pi]$. The power spectrum $d\mu$ is depicted in Figure D.2 and the spectrum has both an absolutely continuous part and a singular part. We would like to identify this spectrum based on covariance estimates and show bounds on the estimation errors. We will use the metric δ_K where $K = \{z \,:\, |z| = 0.9\}$, i.e.,

$$\delta_K(d\mu_0, d\mu_1) = \sup_{|z|=0.9} |P(d\mu_0 - d\mu_1)(z)|.$$

Let **c** be the covariance sequence of $d\mu$ and let $d\mu_5$ and $d\mu_{20}$ be the power spectra with highest entropy in the sets $\mathcal{F}_{\mathbf{c}_{0:5}}$ and $\mathcal{F}_{\mathbf{c}_{0:20}}$, respectively. Figure D.3 compare the spectra $d\mu_5$ and $d\mu$. The first subplot shows the spectra, and the second subplot show $P[d\mu_5](0.9e^{i\theta})$, $P[d\mu](0.9e^{i\theta})$, and bounds on $P[d\nu](0.9e^{i\theta})$ when $\nu \in \mathcal{F}_{\mathbf{c}_{0:5}}$. The spectrum $d\mu_5$ does not distinguish the two peaks and the bounds on the spectra consistent with $\mathbf{c}_{0:5}$ are quite large. For identifying the two spectral lines, the information from $\mathbf{c}_{0:5}$ is clearly not enough, and this data does not provide enough information for the variance to be small in the resolution implied by the metric δ_K.

Figure D.4 compares $d\mu_{20}$ and $d\mu$ similarly. The spectrum $d\mu_{20}$ has two peaks close to the spectral lines and $P[d\mu_{20}](0.9e^{i\theta})$ resembles $P[d\mu](0.9e^{i\theta})$ closely. In fact, as can be seen from the bounds, for any spectrum $\nu \in \mathcal{F}_{\mathbf{c}_{0:20}}$, $P[d\nu](0.9e^{i\theta})$ is close to $P[d\mu](0.9e^{i\theta})$, i.e., that the diameter of the uncertainty set is rather small. The spectra

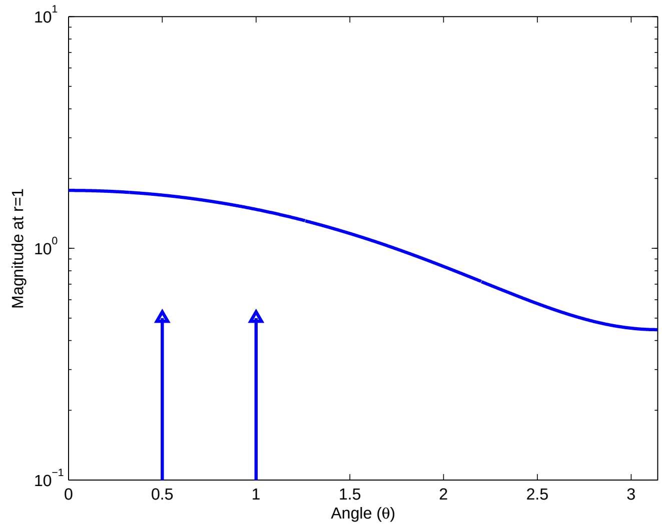

Figure D.2: *The true spectrum* $d\mu$.

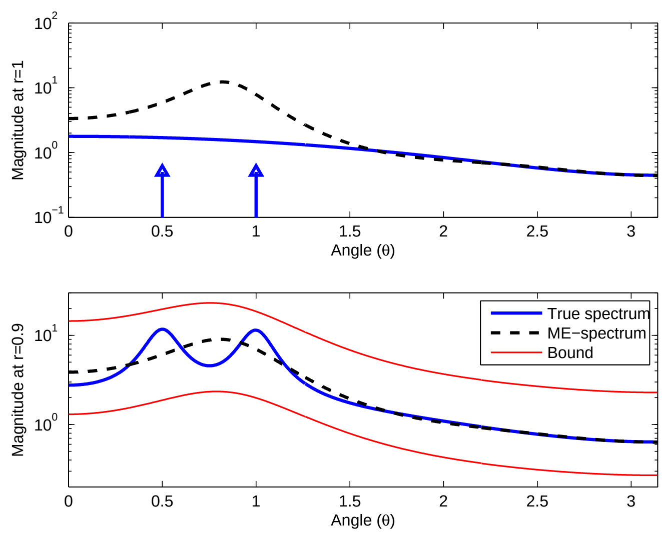

Figure D.3: *The true spectrum* $d\mu$*, the maximum entropy spectrum* $d\mu_5$*, and bounds, based on the covariances* $\mathbf{c}_{0:5}$.

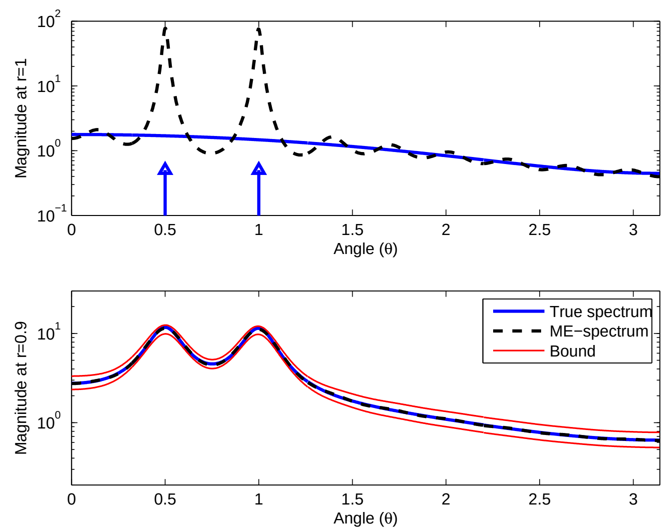

Figure D.4: *The true spectrum $d\mu$, the maximum entropy spectrum $d\mu_{20}$, and bounds, based on the covariances $\mathbf{c}_{0:20}$.*

$d\mu_{\text{det}}$ is the (unique) power spectrum in $\mathcal{F}_{\mathbf{c}_{0:20}}$ with shur parameter $\gamma_{21} = 1$. This is a deterministic spectrum and is depicted in Figure D.5. The figure show that the spectrum respect the bound on $P[d\mu_{\text{det}}](0.9e^{i\theta})$.

The spectra $d\mu$, $d\mu_{20}$, and $d\mu_{\text{det}}$ are very different in terms of their singular and absolutely continuous part. Nevertheless they represent similar distribution of mass and are hence close to each other in the weak*-topology.

D.7 Conclusions and further directions

The choice of metric is essential in every quantitative scientific theory. Identification of power spectra is based on measured covariances and we therefore require that the metric on power spectra is continuous and robust wich respect to uncertainty in the covariances. We show that any such metric need to localize “spectral mass,” or equivalently the metric need to be weak*-continuous. For the metric δ_K, we explicitly quantize the size of the uncertainty set, i.e., the set of all spectra consistent with an covariance estimate. This provides the tools needed for robust identification and where the distances in spectra correspond well to our intuition (see e.g. example 10 in [6]).

There are many directions in which this work could be expanded for further developing the theory. One example of this is approximation in weak*-continuous metrics. It is well known that the Periodogram converges in weak* as the sample size goes to infinity (see e.g., [19]). This makes the Periodogram a suitable candidate for approximations. We

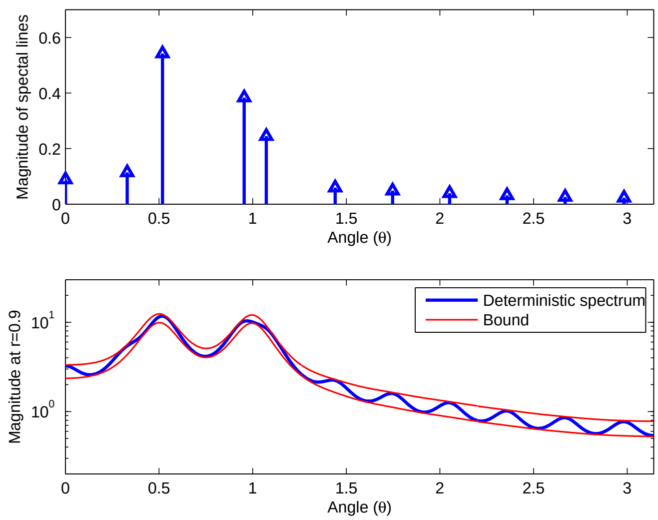

Figure D.5: *The spectrum* $d\mu_{det}$.

expect that the MA-approximation problem can be formulated using standard LMIs. However, optimal ARMA-approximation is more interesting, but also more difficult so solve. This will be subject to further research.

An other direction relate more to the usefullness of the chosen metric for particular applications and is treated in [6]. Here the Monge-Kantorovic transportation distance is exploited in depth, it is weak* and satisfies several additional desirable properties. In particular, it is interesting to study the geodesics of these metrics. It turns out that the geodesics preserve lumpedness of the spectra, and they can therefore model smooth mass transfer [14]. To understand how this preservation of lumpedness relate to the choice of metric will also be studied further.

Appendix

Proof of Theorem D.1. The cannonical neighborhood basis for a point $d\nu$ in the weak* topology on $\mathfrak{M}$ is the sets of the type

$$N(d\nu, \{g_k\}_{k=1}^n, \epsilon) = \left\{ d\mu \geq 0 \,:\, \left| \int_{\mathbb{T}} g_k(d\nu - d\mu) \right| < \epsilon, k = 0, 1, \ldots, n \right\},$$

where g_k are continuous functions on $\mathbb{T}$ for $k = 0, \ldots, n$. To show the theorem we will prove that the neighborhood basis

$$\mathfrak{N}(d\nu) = \{N(d\nu, \{g_k\}_{k=0}^n, \epsilon) \,:\, \epsilon > 0, n \in \mathbb{N}, \{g_k\}_{k=0}^n \subset C(\mathbb{T})\}$$

is equivalent to the neighborhood basis

$$\mathfrak{F}(d\nu) = \left\{ \mathcal{F}_{\mathbf{c}_{0:n},\epsilon} \,:\, \epsilon > 0, n \in \mathbb{N}, c_k = \int_{\mathbb{T}} z^{-k} d\nu, k = 0, \dots, n \right\}.$$

First note that $\mathfrak{N}(d\nu) \supset \mathfrak{F}(d\nu)$, and hence the weak*-topology is at least as strong as the topology induced by $\mathfrak{F}(d\nu)$. To show the other direction, let $N(d\nu, \{g_k\}_{k=0}^n, \epsilon)$ be an arbitrary set in $\mathfrak{N}(d\nu)$. To show the equivalence, it is enough to show that there exists $\epsilon' > 0, n' \in \mathbb{N}$ such that

$$\mathcal{F}_{\mathbf{c}_{0:n'},\epsilon'} \subset N(d\nu, \{g_k\}_{k=1}^n, \epsilon). \tag{D.11}$$

If $\epsilon_k > 0, n_k \in \mathbb{N}$ are such that $\mathcal{F}_{\mathbf{c}_{0:n_k},\epsilon_k} \subset N(d\nu, g_k, \epsilon)$ holds for $k = 1, \dots, n$, then

$$\epsilon' = \min_{k=1,\dots,n} \epsilon_k, \quad n' = \max_{k=1,\dots,n} n_k$$

will satisfy (D.11). We therefore only need to consider the case $n = 1$, and to simplify notation, let $g = g_1$. The function g is continuous and may be approximated uniformly by pseudopolynomials. Let $N' \in \mathbb{N}$ and α_k for $-N' \leq k \leq N$ be such that

$$\left| g(z) - \sum_{\ell=-N'}^{N'} \alpha_\ell z^\ell \right| < \frac{\epsilon}{4\nu(\mathbb{T}) + 2} \text{ for } z \in \mathbb{T},$$

and let $\epsilon' = \min\left(1, \frac{\epsilon}{2\sum|\alpha_\ell|}\right)$. Let $d\mu \in \mathcal{F}_{\mathbf{c}_{0:n'},\epsilon'}$, then

$$\begin{aligned}
\left| \int_{\mathbb{T}} g(d\mu - d\nu) \right| &\leq \int_{\mathbb{T}} \left| g - \sum_{\ell=-N'}^{N'} \alpha_\ell \right| |d\mu - d\nu| + \sum_{\ell=-N'}^{N'} \left|\alpha_\ell z^\ell\right| \left| \int_{\mathbb{T}} (d\mu - d\nu) \right| \\
&< \frac{\epsilon}{4\nu(\mathbb{T}) + 2}(2\nu(\mathbb{T}) + \epsilon') + \sum_{\ell=-N'}^{N'} |\alpha_\ell| \, \epsilon' \\
&\leq \frac{\epsilon}{2} + \frac{\epsilon}{2} = \epsilon,
\end{aligned}$$

and hence $d\mu \in N(d\nu, g, \epsilon)$. We have shown that the topology induced by the neighborhood basis $\mathfrak{F}(d\nu)$ is the weak*-topology, and hence δ is weak*-continuous if and only if (D.2) holds. □

Proof of Proposition D.6. There exists an analytic function $f(z) = H[d\mu](z), d\mu \in \mathcal{F}_{\mathbf{c}_{0:n}}$, such that $f(z) = w_z$ if and only if its associated Pick matrix is nonnegative [17], i.e.

$$\begin{pmatrix} 2T_n & b_z w_z - d_z \\ \bar{w}_z b_z^* - d_z^* & \frac{w_z + \bar{w}_z}{1 - z\bar{z}} \end{pmatrix} \geq 0. \tag{D.12}$$

By using Schur's lemma and completing the squares we arrive at

$$\left| w_z - \frac{\frac{2}{1-z\bar{z}} + (d_z, b_z)}{(b_z, b_z)} \right|^2 \leq \left| \frac{\frac{2}{1-z\bar{z}} + (b_z, d_z)}{(b_z, b_z)} \right|^2 - \frac{(d_z, d_z)}{(b_z, b_z)}, \tag{D.13}$$

where equality holds if and only if the Pick matrix (D.12) is singular. From this, the first part of Proposition D.6 follows. Since the maximum is obtained when equality holds in Equation D.13, the associated Pick matrices are singular. Hence the solutions are unique and correspond to measures with support on $n+1$ points [7, prop. 2]. □

D.8 Bibliography

[1] N. I. Akhiezer. *The classical moment problem*. Hafner Publishing, Translation 1965. Oliver and Boyd.

[2] J. Ball, I. Gohberg, and L. Rodman. *Interpolation of rational matrix functions*, volume 45 of *Operator Theory: Advances and Applications*. Boston, Birkhaüser, 1990.

[3] C. I. Byrnes, T. T. Georgiou, and A. Lindquist. A generalized entropy criterion for Nevanlinna-Pick interpolation with degree constraint. *IEEE Trans. Automat. Control*, 46(6):822–839, June 2001.

[4] J. Capon. High-resolution frequency-wave number spectrum analysis. *IEEE Proc.*, 57:1408–1418, 1969.

[5] C. Foias and A.E. Frazho. *The commutant lifting approach to interpolation problems*. Advances and Applications. Birkhauser Verlag, 1990.

[6] T. T. Georgiou, J. Karlsson, and M. S. Takyar. Metrics for power spectra: an axiomatic approach. Submitted to IEEE Trans on Signal Processing.

[7] T.T. Georgiou. Spectral estimation via selective harmonic amplification: Music, redux. *IEEE Trans. on Signal Proc.*, 48(3):780–790, March 2000.

[8] T.T. Georgiou. Spectral estimation via selective harmonic amplification. *IEEE Trans. on Automatic Contr.*, 46(1):29–42, January 2001.

[9] T.T. Georgiou. Spectral analysis based on the state covariance: the maximum entropy spectrum and linear fractional parameterization. *IEEE Trans. on Automatic Control*, 47(11):1811–1823, November 2002.

[10] Ya. L. Geronimus. *Orthogonal Polynomials*. Consultants Bureau, 1961.

[11] U. Grenander and G. Szegö. *Toeplitz Forms and their Applications*. Chelsea, 1958.

[12] S. Haykin. *Nonlinear Methods of Spectral Analysis*. Springer-Verlag, New York, 1979.

[13] K. Hoffman. *Banach Spaces of Analytic Functions*. Dover Publications, 1962.

[14] X. Jiang, S. Takyar, and T.T. Georgiou. *Metrics and morphing of power spectra*, chapter in Lecture Notes in Control and Information Sciences. Springer Verlag, 2008. to appear.

[15] J. Karlsson and T.T. Georgiou. Signal analysis, moment problems & uncertainty measures. In *IEEE Proc. CDC*, pages 5710– 5715, December 2005.

[16] A. N. Kolmogorov and S. V. Fomin. *Introductory Real Analysis*. Dover Publications, 1970.

[17] I.V. Kovalishina and V.P. Potapov. *Integral Representation of Hermitian Positive Functions*. Khark'hov Railway Engineering Inst., 1981. English translation by T. Ando, Sapporo, Japan.

[18] M.G. Krein and A.A. Nudel'man. *The Markov Moment Problem and Extremal Problems*. American Mathematical Society, Providence, 1977.

[19] E. Parzen. Mathematical considerations in the estimation of spectra. *Technometrics*, 3(2), May 1961.

[20] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

[21] P. P. Vaidyanathan. *Multirate Systems and Filter Banks*. Prentice-Hall, Englewood Cliffs, NJ, 1993.

[22] C. Villani. *Topics in Optimal Transportation*, volume 58. Graduate studies in Mathematics, AMS, 2003.

Metrics for Power Spectra: an Axiomatic Approach

Tryphon Georgiou, Johan Karlsson, and Mir Shahrouz Takyar

Abstract

We present an axiomatic framework for seeking distances between power spectral density functions. The axioms require that the sought metric respects the effects of additive and multiplicative noise in reducing our ability to discriminate spectra, as well as they require continuity of statistical quantities with respect to perturbations measured in the metric. We then present a particular metric which abides by these requirements. The metric is based on the Monge-Kantorovich transportation problem and is contrasted with an earlier Riemannian metric based on the minimum-variance prediction geometry of the underlying time-series. It is also being compared with the more traditional Itakura-Saito distance measure, as well as the aforementioned prediction metric, on two representative examples.

E.1 Introduction

A key element of any quantitative scientific theory is a well-defined and natural metric. A model for the development of such metrics is provided, in the context of information theory and statistics, in the work of Fisher, Rao, Amari, Centsov and many others, via an axiomatic approach where the sought metric is identified on the basis of a natural set of axioms—the main one being the contractiveness of stochastic maps. The subject of the present paper is not the geometry of information, but instead, the possibility of analogous geometries for power spectra starting from a similar axiomatic rationale. Specifically, we seek a metric between power spectra which is contractive when noise is introduced, since intuitively, noise impedes our ability to discriminate. Further, we require that statistics are continuous with respect to spectral uncertainty quantified by the sought metric. We build on [17] where a variety of potential metrics were studied using complex analysis. The focus of the current paper is twofold, firstly to propose a natural set of axioms that geometries for power spectra must satisfy, and secondly to present a particular candidate which abides by the stated axioms. This metric is based on the Monge-Kantorovich transportation problem and represents a relaxation of Wasserstein distances so as to be applicable to power spectra.

In Section E.2 we outline and discuss the axiomatic framework. In Section E.3 we contrast the present setting with two alternatives, first the axiomatic basis of Information Geometry and then with a geometry that is inherited by linear prediction theory. In Section E.4 we present basic facts of the Monge-Katnorovich transportation problem which are then utilized in Section E.5 in order to develop a suitable family of metrics satisfying the axioms of the sought spectral geometry.

E.2 Morphisms on power spectra

We consider power spectra of discrete-time stochastic processes. These are bounded positive measures on the interval $\mathbb{I} = [-\pi, \pi]$ (with the end points identified) or, in the case of real-valued processes, on $\mathbb{I} = [0, \pi]$ and the set of such measures is denoted by

$$\mathfrak{M} := \{d\mu \, : \, d\mu \geq 0 \text{ on } \mathbb{I}\} \, .$$

The physics of signal interactions suggests certain natural morphisms between spectra that model mixing in the time-domain. The most basic such interactions, additive and multiplicative, adversely affect the information content of signals. It is our aim to devise metrics that respect such a degradation in information content. Another property that ought to be inherent in a metric geometry for power spectra is the continuity of statistics. More specifically, since modeling and identification is often based on statistical quantities, it is natural to demand that "small' changes in the spectral content, as measured by suitable metrics, result in small changes in any relevant statistical quantity.

Consider a discrete-time stationary (in general complex-valued) zero-mean stochastic process $\{y(k), \ k \in \mathbb{Z}\}$, or simply y for short, with corresponding power spectrum $d\mu \in \mathfrak{M}$. The sequence of covariances

$$R(\ell) := \mathcal{E}\{y(m)\overline{y(m-\ell)}\}, \ \ell = 0, 1, 2, \ldots,$$

where $\mathcal{E}\{\cdot\}$ denotes expectation and " $^{-}$ " denotes complex conjugation, are the Fourier coefficients of $d\mu$, i.e.,

$$R(\ell) = \int_{\mathbb{I}} e^{-j\ell\theta} d\mu(\theta).$$

In general, second order statistics are integrals of the form

$$\mathbf{R} = \int_{\mathbb{I}} \mathbf{G}(\theta) d\mu(\theta)$$

for an arbitrary vectorial integration kernel $\mathbf{G}(\theta)$ which is continuous in $\theta \in \mathbb{R}$ and periodic with period 2π. For future reference we denote the set of such functions by $\mathcal{C}_{perio}(\mathbb{I})$.

Now, suppose that $d\mu_a$ represents the power spectrum of an "additive-noise" process y_a which is independent of y. Then the power spectrum of $y + y_a$ is simply $d\mu + d\mu_a$.

Similarly, if $d\mu_m$ represents the power spectrum of a "multiplicative-noise" process y_m, the power spectrum of the point-wise product $y \cdot y_m$ is the circular convolution

$$d\nu = d\mu * d\mu_m.$$

I.e., $d\nu$ satisfies

$$\int_{x\in S} d\nu(x) := \int_{x\in S}\int_{t\in\mathbb{I}} d\mu(t)d\mu_a(x-t) \text{ for all } S\subseteq\mathbb{I},$$

where the arguments are interpreted modulo 2π.

We postulate situations where we need to discriminate between two signals on the basis of their power spectra and of their statistics. In such cases, additive or multiplicative noise may impede our ability to differentiate between the two. Thus, we consider noise spectra as morphisms on $\mathfrak{M}$ that transform power spectra accordingly. Thus, additive and multiplicative noise morphisms are defined as follows:

$$A_{d\mu_{\rm a}} \ : \ d\mu \mapsto d\mu + d\mu_{\rm a}$$

for any $d\mu_a \in \mathfrak{M}$, and

$$M_{d\mu_{\rm m}} \ : \ d\mu \mapsto d\mu * d\mu_{\rm m}$$

for any $d\mu_{\rm m} \in \mathfrak{M}$, normalized so that $\int_{\mathbb{I}} d\mu_{\rm m} = 1$. The normalization is such that multiplicative noise is perceived to affect the spectral content but not the total energy of underlying signals.

The effect of additive independent noise on the statistics of a process is also additive, e.g., covariances of the process are transformed according to

$$\hat{A}_{d\mu_{\rm a}} \ : \ R(\ell) \mapsto R(\ell) + R_{\rm a}(\ell),$$

where $R_{\rm a}(\ell)$ denotes the corresponding covariances of the noise process. Similarly, multiplicative noise transforms the process statistics by pointwise multiplication (Schur product) as follows

$$\hat{M}_{d\mu_{\rm m}} \ : \ R(\ell) \mapsto R(\ell) \cdot R_{\rm m}(\ell).$$

More generally, $\hat{M}_{d\mu_{\rm m}} \ : \ \mathbf{R} \mapsto \mathbf{R} \bullet \mathbf{R}_{\rm m}$ for statistics with respect to an arbitrary kernel $\mathbf{G}(\theta)$, where $\bullet$ denotes point-wise multiplication of the vectors $\mathbf{R}$, $\mathbf{R}_m$.

Consistent with the intuition that noise masks differences between two power spectra, it is reasonable to seek a metric topology, where distances between power spectra are non-increasing when they are transformed by any of the above two morphisms. More precisely, we seek a notion of distance $\delta(\cdot,\cdot)$ on $\mathfrak{M}$ with the following properties:

Axiom i) $\delta(\cdot,\cdot)$ is a metric on $\mathfrak{M}$.

Axiom ii) For any $d\mu_a \in \mathfrak{M}$, $A_{d\mu_{\rm a}}$ is contractive on $\mathfrak{M}$ with respect to the metric $\delta(\cdot,\cdot)$.

Axiom iii) For any $d\mu_m \in \mathfrak{M}$ with $\int_{\mathbb{I}} d\mu_{\rm m} \leq 1$, $M_{d\mu_{\rm m}}$ is contractive on $\mathfrak{M}$ with respect to the metric $\delta(\cdot,\cdot)$.

The property of a map being contractive refers to the requirement that the distance between two power spectra does not increase when the transformation is applied.

An important property for the sought topology of power spectra is that small changes in the power spectra are reflected in correspondingly small changes in statistics. More precisely, any topology induces a notion of convergence, and the question is whether this topology is compatible with the topology of the Euclidean vector-space where (finite) statistics take their values. Continuity of statistics to changes in the power spectra is necessary for quantifying spectral uncertainty based on statistics. The property we require is referred to as weak*-continuity and is abstracted in the following statement.

Axiom iv) Let $d\mu \in \mathfrak{M}$ and a sequence $d\mu_k \in \mathfrak{M}$ for $k \in \mathbb{N}$. Then $\delta(d\mu_k, d\mu) \to 0$ as $k \to \infty$, if and only if

$$\int_{\mathbb{I}} \mathbf{G} d\mu_k \to \int_{\mathbb{I}} \mathbf{G} d\mu \text{ as } k \to \infty,$$

for any $\mathbf{G} \in \mathcal{C}_{perio}(\mathbb{I})$.

E.3 Reflections and contrast with information geometry

The search for natural metrics between density functions can be traced back to the early days of statistics, probability and information theory. According to Chentsov [8, page 992] (ref. [1]), Kolmogorov was "always interested in finding *information* distances" between probability distributions. In his notes he emphasized the importance of the total variation

$$d_{\mathrm{TV}}(d\mu_0, d\mu_1) := \int |\mu_0(dx) - \mu_1(dx)|$$

as a metric, and he independently arrived at and discussed the relevance of the Bhattacharyya [5] distance

$$d_{\mathrm{B}}(d\mu_0, d\mu_1) := 1 - \int \sqrt{\mu_0(dx)\mu_1(dx)} \tag{E.1}$$

as a measure of unlikeness of two measures $d\mu_0, d\mu_1$. Both suggestions reveal great intuition and foresight. The total variation admits the following interpretation (cf. [11]) that will turn out to be particularly relevant in our context: the total variation represents the least "energy" of perturbations of two power spectra $d\mu_0$ and $d\mu_1$ that render the two indistinguishable, i.e.,

$$\begin{aligned} d_{\mathrm{TV}}(d\mu_0, d\mu_1) \quad = \quad & \min\{\int d\nu_0 + \int d\nu_1 \ : \\ & d\nu_0,\, d\nu_1 \in \mathfrak{M},\ \text{and } d\mu_0 + d\nu_0 = d\mu_1 + d\nu_1\}. \end{aligned} \tag{E.2}$$

On the other hand the Bhattacharyya distance turned out to have deep connections with Fisher information, the Kullback-Leibler divergence, and the Cramér-Rao inequality. These connections underlie a body of work known as Information Geometry which was advanced by Amari, Nagaoka, Chentsov and others [18, 7, 2]. At the heart of the subject is the

Fisher information metric on probability spaces and the closely related spherical Fisher-Bhattacharyya-Rao metric

$$d_{\text{FBR}}(d\mu_0, d\mu_1) := \arccos \int \sqrt{\mu_0(dx)\mu_1(dx)}. \tag{E.3}$$

This latter metric is precisely the geodesic distance between two distributions in the geometry of the Fisher metric. One of the fundamental results of the subject is Chentsov's theorem. This theorem states that stochastic maps are contractive with respect to the Fisher information metric and moreover, that this metric is in fact the *unique* (up to constant multiple) Riemannian metric with this property [7]. Stochastic maps represent the most general class of linear maps which map probability distributions to the same. Stochastic maps model coarse graining of the outcome of sampling, and thus, form a semi-group. Thus, it is natural to require that any natural notion of distance between probability distributions must be monotonic with respect to the action of stochastic maps.

An alternative justification for the Fisher information metric is based on the Kullback-Leibler divergence

$$d_{\text{KL}}(d\mu_0, d\mu_1) := \int \frac{d\mu_0}{d\mu_1} \log(\frac{d\mu_0}{d\mu_1}) d\mu_1 = \int \log(\frac{d\mu_0}{d\mu_1}) d\mu_0$$

between *probability* distributions. The Kullback-Leibler divergence is not a metric, but quantifies in a very precise sense the difficulty in distinguishing the two distributions [20]. In fact, it may be seen to quantify, in source coding for discrete finite probability distributions, the increase in the average word-length when a code is optimized for one distribution and used instead for encoding symbols generated according to the other. The distance between infinitesimal perturbations, measured using d_{KL}, is precisely the Fisher information metric. It is quite remarkable that both lines of reasoning, degradation of coding efficiency and ability to discriminate on one hand and contractiveness of stochastic maps on the other, lead to the same geometry on probability spaces.

Turning again to power spectra, we observe that d_{TV} can be used as a metric and has a natural interpretation as explained earlier. The metric d_{FBR} on the other hand can also be used, if suitably modified to account for scaling, but lacks an intrinsic interpretation. A variety of other metrics can also be placed on $\mathfrak{M}$ (cf. [14, 16, 21]), mostly borrowed from functional analysis, which may similarly lack an intrinsic interpretation. Thus, the signal processing community focused instead on other metric-like quantities, as the so-called Itakura-Saito distance [14, Equation (16)]

$$d_{\text{IS}}(f_0 d\theta, f_1 d\theta) := \int \left(\frac{f_0}{f_1} - \ln(\frac{f_0}{f_1}) - 1 \right) d\theta,$$

which have been motivated in the context of linear prediction [22, 14]. This relates is intimately related to the probability structure of underlying processes for the Gaussian case and to their distance in the Kullback-Leibler divergence (see for instance [19, 23]). A closely related metric was presented in [10, 12] which quantifies in a precise way the degradation of predictive error variance –in analogy with the latter argument that led to the Fisher metric. More specifically, a one-step optimal linear predictor for an underlying random process is obtained based on a given power spectrum, and then, this predictor is applied to a random process with a different spectrum. The degradation of predictive error variance, when the perturbations are infinitesimal, gives rise to a Riemannian metric. In this metric, the geodesic distance between two power spectra is

$$d_{\text{pr}}(d\mu_0, d\mu_1) := \sqrt{\int (\log \frac{d\mu_0}{d\mu_1})^2 d\theta - (\int \log \frac{d\mu_0}{d\mu_1} d\theta)^2}, \tag{E.4}$$

which effectively depends on the ratio of the corresponding spectral densities. Interestingly, this metric is a "normalized version" of the so-called log-spectral distance [21, Equation (78)]

$$d_{\text{LS}}(d\mu_0, d\mu_1) := \sqrt{\int (\log \frac{d\mu_0}{d\mu_1})^2 d\theta}$$

which is commonly used without any intrinsic justification. A similar rationale that lead to (E.4) can be based on degradation of smoothing-variance instead of prediction (see [10, 12]), and this also leads to expressions that weigh in ratios of the corresponding spectral density functions, i.e., it is the ratios of the absolutely continuous part of the measures that play any role.

A possible justification for such metrics which weigh in only the ratio of the corresponding density functions can be sought in interpreting the effect of linear filtering as a kind of processing that needs to be addressed in the axioms. More specifically, the power spectrum at the output of a linear filter relates to the power spectrum of the input via multiplication by the modulus square of the transfer function. Thus, a metric that respects such "processing" ought to be contractive (and possibly invariant). However, it turns out that such a property is incompatible with the spectral properties that we would like to have, and in particular it is incompatible with the ability of the metric to localize a measure based on its statistics (cf. Axiom iv). This incompatibility is shown next.

Consider morphisms on $\mathfrak{M}$ that correspond to processing by a linear filter:

$$F_h \ : \ d\mu \mapsto |h|^2 d\mu$$

for any $h \in H_\infty$. Here, h is thought of as the transfer function of the filter, μ the power spectrum of the input, and $|h|^2 d\mu$ the power spectrum of the output.

Proposition E.1. *Assume that $\delta(\cdot,\cdot)$ is a weak*-continuous metric on $\mathfrak{M}$. Then there exists $h \in H_\infty$ such that F_h is not contractive with respect to $\delta(\cdot,\cdot)$.*

Proof. We will prove the claim by showing that whenever δ is a weak*-continuous metric that is a contraction with respect to F_h, we may derive a contradiction. Denote by μ_t, $t \geq 0$ the measure with a unit mass in the point t and let $\epsilon = \delta(d\mu_0, d\mu_0/2)$. By weak*-continuity, there exists $t_0 > 0$ such that $\delta(d\mu_0, d\mu_{t_0}) < \epsilon/3$. Let $h \in H_\infty$ be such that

$|h(0)|^2 = 1/2$ and $|h(t_0)|^2 = 1$. Then we have that

$$\begin{aligned} \epsilon = \delta(d\mu_0, d\mu_0/2) &\leq \delta(d\mu_0, d\mu_{t_0}) + \delta(d\mu_{t_0}, d\mu_0/2) \\ &= \delta(d\mu_0, d\mu_{t_0}) + \delta(|h|^2\mu_{t_0}, |h|^2\mu_0) \\ &\leq 2\delta(d\mu_0, d\mu_{t_0}) < \frac{2}{3}\epsilon. \end{aligned}$$

Which is a contradiction, and hence the proposition holds. □

It is important to point out that none of the above (i.e., neither (E.3) nor (E.4)) is a weak*-continuous metric. In particular, the metric in (E.4) is impervious to spectral lines as only the absolutely continuous part of the spectra plays any role. Similarly, neither the metric in (E.2) nor the one in (E.3) can localize distributions because they are not weak*-continuous. Thus, in this paper, we follow a line of reasoning analogous to the axiomatic framework of the Chentsov theorem, but for power spectra, requiring a metric to satisfy Axioms i)-iv).

E.4 The Monge-Kantorovich problem

A natural class of metrics on measures are transport metrics based on the ideas of Monge and Kantorovich. The Monge-Kantorovich distance represents a cost of moving a nonnegative measure $d\mu_0 \in \mathfrak{M}(X)$ to another nonnegative measure $d\mu_1 \in \mathfrak{M}(X)$, given that there is an associated cost $c(x, y)$ of moving mass from the point x to the point y. The theory may be formulated for rather general spaces X, but in this paper we restrict our attention to compact metric spaces X.Every possible way of moving the measure $d\mu_0$ to $d\mu_1$ corresponds to a transference plan $\pi \in \mathfrak{M}(X \times X)$, which satisfies

$$\int_{y\in X} d\pi(x, y) = d\mu_0 \text{ and } \int_{x\in X} d\pi(x, y) = d\mu_1,$$

or more rigorously, that

$$\pi[A \times X] = \mu_0(A) \text{ and } \pi[X \times B] = \mu_1(B) \tag{E.5}$$

whenever $A, B \subset X$ are measurable. Such a plan exists only if the measures $d\mu_0$ and $d\mu_1$ have the same mass, i.e. $\mu_0(X) = \mu_1(X)$. Denote by $\Pi(d\mu_0, d\mu_1)$ the set of all such transference plans, i.e.

$$\Pi(d\mu_0, d\mu_1) = \{\pi \in \mathfrak{M}(X \times X) \,:\, \text{(E.5) holds for all } A, B\}.$$

To each such transference plan, the associated cost is

$$\mathcal{I}[\pi] = \int_{X\times X} c(x, y) d\pi(x, y)$$

and consequently, the minimal transportation cost is

$$T_c(d\mu_0, d\mu_1) := \min\{\mathcal{I}(\pi) \,:\, \pi \in \Pi(d\mu_0, d\mu_1)\}. \tag{E.6}$$

The optimal transportation problem admits a dual formulation, referred to as the Kantorovich duality, which we state below without a proof. For a derivation and an insightful exposition we refer the reader to [25, page 19].

Theorem E.2. *Let c be a lower semi-continous (cost) function, let*

$$\Phi_c := \{(\phi, \psi) \in L^1(d\mu_0) \times L^1(d\mu_1) : \phi(x) + \psi(y) \le c(x, y)\},$$

and let

$$\mathcal{J}(\phi, \psi) = \int_X \phi d\mu_0 + \psi d\mu_1.$$

Then

$$T_c(d\mu_0, d\mu_1) = \sup_{(\phi,\psi)\in\Phi_c} \mathcal{J}(\phi, \psi).$$

A rather simple consequence is the following lemma.

Lemma E.3. *Let c be a lower semi-continous (cost) function with $c(x, x) = 0$ for $x \in X$. Then $A_{d\mu_a}$ is contractive with respect to T_c.*

Proof. Contractiveness of $A_{d\mu_{\mathrm{a}}}$ follows from the dual representation. Any pair $(\phi, \psi) \in \Phi_c$ satisfies $\phi(x) + \psi(x) \le 0$, and hence

$$\int_X \phi d\mu_0 + \psi d\mu_1 \ge \int_X \phi d\mu_0 + \psi d\mu_1 + (\phi + \psi) d\mu_{\mathrm{a}}.$$

□

Monge-Kantorovich distances are not metrics, in general, but they readily give rise to a class of the so-called Wasserstein metrics as explained next.

Theorem E.4. *Assume that the (cost) function $c(\cdot, \cdot)$ is of the form $c(x, y) = d(x, y)^p$ where d is a metric and $p \in (0, \infty)$. Then the Wasserstein distance*

$$W_p(d\mu_0, d\mu_1) = T_c(d\mu_0, d\mu_1)^{\min(1, \frac{1}{p})}$$

is a metric on the subspace of $\mathfrak{M}(X)$ with fixed mass and metrizes the weak* topology.

Proof. See [25, Chapter 7]. Note that since X is compact, the weak* topology on $\mathfrak{M}(X)$ coincides with the weak topology. □

E.5 Metrics based on transportation

The Monge-Kantorovich theory deals with measures of equal mass. As we have just seen, it provides metrics that have some of the properties that we seek to satisfy. The purpose of this section is to develop a metric based on similar principles, that applies to measures of possibly unequal mass.

Given two nonnegative measures $d\mu_0$ and $d\mu_1$ on $\mathbb{I}$, we postulate that these are perturbations of two other measures $d\nu_0$ and $d\nu_1$, respectively, which have equal mass. Then, the cost of transporting $d\mu_0$ and $d\mu_1$ to one another can be thought of as the cost of transporting $d\nu_0$ and $d\nu_1$ to one another plus the size of the respective perturbations. Thus we define

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) := \inf_{\nu_0(\mathbb{I})=\nu_1(\mathbb{I})} T_c(d\nu_0, d\nu_1) + \kappa \sum_{i=0}^{1} d_{\mathrm{TV}}(d\mu_i, d\nu_i), \tag{E.7}$$

where κ is a suitable parameter that weighs the relative contribution of perturbation and transportation. Define

$$c(x, y) = |(x - y)_{\mathrm{mod}2\pi}|^p \tag{E.8}$$

where $(x)_{\mathrm{mod}2\pi}$ is the element in the equivalence class $x + 2\pi\mathbb{Z}$ which belongs to $(-\pi, \pi]$. The main result of the section is the following theorem.

Theorem E.5. *Let $\kappa > 0$ and $c(x, y)$ defined as in* (E.8)*, where $p \in (0, \infty)$. Then*

$$\delta_{p,\kappa}(d\mu_0, d\mu_1) := \left(\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)\right)^{\min(1,\frac{1}{p})}$$

is a metric on $\mathfrak{M}$ which satisfies Axiom i) - iv).

The proof uses the fact that (E.7) has an equivalent formulation as a transportation problem, and a corresponding dual stated below.

Theorem E.6. *Let c be a lower semi-continuous (cost) function, let*

$$\begin{aligned}\Phi_{c,\kappa} \quad := \quad & \big\{(\phi, \psi) \in L^1(d\mu_0) \times L^1(d\mu_1) \,:\, \phi(x) \le \kappa, \\ & \psi(y) \le \kappa\,,\ \phi(x) + \psi(y) \le c(x, y)\big\},\end{aligned}$$

and let

$$\mathcal{J}(\phi, \psi) = \int_{\mathbb{I}} \phi d\mu_0 + \psi d\mu_1.$$

Then

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = \sup_{(\phi,\psi)\in\Phi_{c,\kappa}} \mathcal{J}(\phi, \psi). \tag{E.9}$$

Remark E.1. Definition (E.7) does not provide a direct way to compute $\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)$, whereas the dual formulation in Theorem E.6 is amenable to numerical implementation. Indeed, (E.9) is a linear optimization problem which can be computed using standard methods.

Proof. The problem (E.7) can be thought of as a transportation problem on the set $X = \mathbb{I} \cup \{\infty\}$, where a mass is added at ∞ as needed to normalize the measures so that they have equal mass, e.g.,

$$\begin{aligned}\hat{\mu}_i(S) &= \mu_i(S) \text{ for } S \subset \mathbb{I} \\ \hat{\mu}_i(\infty) &= M - \mu_i(\mathbb{I})\end{aligned}$$

for some $M \geq \max\{\mu_i(\mathbb{I}) : i = 0, 1\}$. Accordingly, the (cost) function is modified as follows

$$\hat{c}(x,y) = \begin{cases} \min(c(x,y), 2\kappa) & \text{for} \quad x, y \in \mathbb{I}, \\ \kappa & \text{for} \quad x \in \mathbb{I}, y = \infty, \\ \kappa & \text{for} \quad x = \infty, y \in \mathbb{I}, \\ 0 & \text{for} \quad x = \infty, y = \infty. \end{cases} \tag{E.10}$$

First we prove that

$$T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1) = \sup_{(\phi,\psi) \in \Phi_{c,\kappa}} \mathcal{J}(\phi, \psi),$$

hence $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1)$ is independent of M, and then we conclude the proof by showing that

$$\tilde{T}_{c,\kappa}(d\mu_0, d\mu_1) = T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1). \tag{E.11}$$

According to Theorem E.2, $T_{\hat{c}}(d\hat{\mu}_0, d\hat{\mu}_1)$ is equal to the supremum of

$$\hat{\mathcal{J}}(\phi, \psi) := \int_X \phi d\hat{\mu}_0 + \psi d\hat{\mu}_1$$

subject to

$$\phi(x) + \psi(y) \leq \hat{c}(x,y) \quad \text{for} \quad x, y \in \mathbb{I}, \tag{E.12}$$

$$\phi(x) + \psi(\infty) \leq \kappa \quad \text{for} \quad x \in \mathbb{I}, \tag{E.13}$$

$$\phi(\infty) + \psi(y) \leq \kappa \quad \text{for} \quad y \in \mathbb{I}, \tag{E.14}$$

$$\phi(\infty) + \psi(\infty) \leq 0. \tag{E.15}$$

Our first claim now follows by showing that there is no added restriction imposed by requiring that $\phi(\infty) = \psi(\infty) = 0$. Indeed, $\Phi_{c,\kappa}$ is essentially identical to the set

$$\{(\phi, \psi) : \text{(E.12)} - \text{(E.15) hold and } \phi(\infty) = \psi(\infty) = 0\},$$

with (ϕ, ψ) extended to have support at ∞ as well. To this end, let (ϕ, ψ) be an arbitrary pair of functions satisfying (E.12)-(E.15). Since additive scaling of $(\phi, -\psi)$ does not change the constraints nor the value of $\hat{\mathcal{J}}(\phi, \psi)$, we may assume that $\phi(\infty) = 0$. There are two cases that we need to consider. If $\sup_{x \in \mathbb{I}} \phi(x) \leq \kappa$, then define

$$\hat{\phi}(x) = \phi(x) \text{ for } x \in X, \ \hat{\psi}(x) = \begin{cases} \psi(x) & x \in \mathbb{I} \\ 0 & x = \infty, \end{cases}$$

and if $\sup_{x \in \mathbb{I}} \phi(x) = \epsilon + \kappa > \kappa$, then define

$$\hat{\phi}(x) = \begin{cases} \phi(x) - \epsilon & x \in \mathbb{I} \\ 0 & x = \infty \end{cases}, \ \hat{\psi}(x) = \begin{cases} \psi(x) + \epsilon & x \in \mathbb{I} \\ 0 & x = \infty. \end{cases}$$

In both cases we have that $\hat{\mathcal{J}}(\phi, \psi) \leq \hat{\mathcal{J}}(\hat{\phi}, \hat{\psi})$ as well as that the pair $(\hat{\phi}, \hat{\psi})$ satisfies (E.12)-(E.15). In the second case, the constraint (E.14) is not violated; $\hat{c}(x,y) \leq 2\kappa$

implies that $\sup_{x\in\mathbb{I}}\phi(x)+\sup_{y\in\mathbb{I}}\psi(y)\leq 2\kappa$, and hence $\sup_{y\in\mathbb{I}}\psi(y)\leq\kappa-\epsilon$. Note that (E.13) implies that $\psi(\infty)\leq-\epsilon$. Thus, in both cases, from an arbitrary pair (ϕ,ψ), we have constructed a pair $(\hat{\phi},\hat{\psi})$ for which the constraints (E.12)-(E.15) hold, $\hat{\phi}(\infty)=\hat{\psi}(\infty)=0$ holds, and the value of $\hat{\mathcal{J}}(\phi,\psi)$ has not decreased. Therefore we may without loss of generality let $\phi(\infty)=\psi(\infty)=0$.

It remains to show that $\tilde{T}_{c,\kappa}(d\mu_0,d\mu_1) = T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1)$. We start by showing that $T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1)\geq\tilde{T}_{c,\kappa}(d\mu_0,d\mu_1)$. Let $\hat{\pi}\in\Pi(d\hat{\mu}_0,d\hat{\mu}_1)$ and let

$$\int_{y\in\mathbb{I}} d\hat{\pi}(x,y)=d\nu_0 \text{ and } \int_{x\in\mathbb{I}} d\hat{\pi}(x,y)=d\nu_1.$$

Then for any transference plan $\hat{\pi}$ we have that

$$\begin{aligned}
\mathcal{I}[\hat{\pi}] &= \int_{\mathbb{I}\times\mathbb{I}}\hat{c}(x,y)\hat{\pi}(x,y)+\kappa\int_{\mathbb{I}}d\hat{\pi}(x,\infty)+\kappa\int_{\mathbb{I}}d\hat{\pi}(\infty,y)\\
&= \int_{\mathbb{I}\times\mathbb{I}}\hat{c}(x,y)\hat{\pi}(x,y)+\kappa\sum_{i=0}^{1}\int_{\mathbb{I}}(d\mu_i-d\nu_i)\\
&\geq T_c(d\nu_0,d\nu_1)+\kappa\sum_{i=0}^{1}d_{\mathrm{TV}}(d\mu_i,d\nu_i)\\
&\geq \tilde{T}_{c,\kappa}(d\mu_0,d\mu_1),
\end{aligned}$$

by noting that $d\mu_i-d\nu_i$ is positive and hence $\int_{\mathbb{I}}(d\mu_i-d\nu_i)=d_{\mathrm{TV}}(d\mu_i,d\nu_i)$. Therefore $T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1)\geq\tilde{T}_{c,\kappa}(d\mu_0,d\mu_1)$ always holds. To show that the reverse inequality also holds let $d\nu_0,d\nu_1$ be two non-negative measures with $\nu_0(\mathbb{I})=\nu_1(\mathbb{I})$. By introducing $f_0=\max(-\kappa,\phi)$, $f_1=\max(-\kappa,\psi)$ and using the dual formulation we get

$$\begin{aligned}
T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1) &= \sup_{(\phi,\psi)\in\Phi_{c,\kappa}}\mathcal{J}(\phi,\psi)\\
&\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}}\int_{\mathbb{I}}f_0d\mu_0+f_1d\mu_1\\
&\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}}\sum_{i=0}^{1}\int_{\mathbb{I}}f_id\nu_i+f_i(d\mu_i-d\nu_i)\\
&\leq \sup_{(\phi,\psi)\in\Phi_{c,\kappa}}\sum_{i=0}^{1}\int_{\mathbb{I}}f_id\nu_i+\kappa\sum_{i=0}^{1}d_{\mathrm{TV}}(d\mu_i,d\nu_i)\\
&\leq T_c(d\nu_0,d\nu_1)+\kappa\sum_{i=0}^{1}d_{\mathrm{TV}}(d\mu_i,d\nu_i).
\end{aligned}$$

Here we use the fact that $\kappa d_{\mathrm{TV}}(d\mu_i,d\nu_i)=\sup_{\|f\|_\infty\leq\kappa}\int f(d\mu_i-d\nu_i)$. Since the above inequality holds for any measures $d\nu_0,d\nu_1$ we get the reversed inequality $T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1)\leq\tilde{T}_{c,\kappa}(d\mu_0,d\mu_1)$, and hence we conclude that $T_{\hat{c}}(d\hat{\mu}_0,d\hat{\mu}_1)=\tilde{T}_{c,\kappa}(d\mu_0,d\mu_1)$. □

The final step to proving Theorem E.5 will be provided by the following lemma.

Lemma E.7. *Let $c(x,y)$ be a function of $|x-y|$. Then for any $d\mu_m \in \mathfrak{M}$ with $\int_{\mathbb{I}} d\mu_m \leq 1$, $M_{d\mu_m}$ is contractive on $\mathfrak{M}$ with respect to $\tilde{T}_{c,\kappa}$.*

Proof. Note that

$$\begin{aligned}
&\int_{x\in\mathbb{I}} \phi(x)(d\mu_m * d\mu_0)(x) \\
&\qquad = \int_{x\in\mathbb{I}} \phi(x) \int_{\tau\in\mathbb{I}} d\mu_m(x-\tau)d\mu_0(\tau) \\
&\qquad = \int_{\tau\in\mathbb{I}} \left(\int_{x\in\mathbb{I}} \phi(x)d\mu_m(x-\tau) \right) d\mu_0(\tau) \\
&\qquad = \int_{\tau\in\mathbb{I}} \left(\phi(x) * d\mu_m(-x)\right)\big|_\tau \, d\mu_0(\tau),
\end{aligned}$$

and denote

$$\begin{aligned}
\phi_m(\tau) &= \phi(x) * d\mu_m(-x)\big|_\tau \\
\psi_m(\tau) &= \psi(x) * d\mu_m(-x)\big|_\tau \, .
\end{aligned}$$

From this it follows that

$$\mathcal{J}_{(d\mu_m * d\mu_0, d\mu_m * d\mu_1)}(\phi,\psi) = \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi_m, \psi_m),$$

where the subscript specifies the measures used in the definition of the dual functional.

Now let $(\phi,\psi) \in \Phi_{c,\kappa}$. Then

$$\begin{aligned}
\phi(x-\tau) + \psi(y-\tau) &\leq \min\left(c(x-\tau, y-\tau), 2\kappa\right) \\
&= \min\left(c(x,y), 2\kappa\right),
\end{aligned}$$

and by integrating with respect to $d\mu_m(-\tau)$ over $\tau \in \mathbb{I}$, we arrive at

$$\phi_m(x) + \psi_m(y) \leq \min(c(x,y), 2\kappa).$$

Furthermore, it is immediate that $\phi(x) \leq \kappa$ and $\psi(y) \leq \kappa$ implies that $\phi_m(x) \leq \kappa$ and that $\psi_m(y) \leq \kappa$, and hence $(\phi_m, \psi_m) \in \Phi_{c,\kappa}$ follows. Finally

$$\begin{aligned}
&\tilde{T}_{c,\kappa}(M_{d\mu_m}(d\mu_0), M_{d\mu_m}(d\mu_1)) \\
&\qquad = \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_m * d\mu_0, d\mu_m * d\mu_1)}(\phi,\psi) \\
&\qquad = \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi_m,\psi_m) \\
&\qquad \leq \sup_{(\phi,\psi)\in\Psi_{c,\kappa}} \mathcal{J}_{(d\mu_0, d\mu_1)}(\phi,\psi) \\
&\qquad = \tilde{T}_{c,\kappa}(d\mu_0, d\mu_1)
\end{aligned}$$

□

We now recap the proof of our main theorem.

Proof. [Proof of Theorem E.5]: From the formulation (E.11), $\tilde{T}_{c,\kappa}$ can be viewed as the cost of a transportation problem. Since the associated cost function $\hat{c}$ from (E.10) is of the form d^p, where d is a metric, Axioms i) follows from Theorem E.4. Axiom iv) follows by noting that if a sequence of measures $d\mu_n$ converges to $d\nu$ in weak*, then in particular $\mu_n(\mathbb{I}) \to \nu(\mathbb{I})$. Therefore, for any $M > \sup\{\mu_n(\mathbb{I}),\ n \in \mathbb{N}\}$, it readily follows that $\hat{\mu}_n$ converges to $\hat{\nu}$ in weak*, and hence Theorem E.4 ensures weak* continuity of $\delta_{p,\kappa}$. From the above formulation, Axiom ii) follows from Lemma E.3. Finally Axiom iii) follows from Lemma E.7. ☐

Remark E.2. It is interesting to note that for the case $p = 1$

$$\delta_{1,\kappa}(d\mu_0, d\mu_1) = \max_{\substack{\|g\| \leq \kappa \\ \|g\|_L \leq 1}} \int g(d\mu_0 - d\mu_1),$$

where $\|f\|_L = \sup \frac{|f(x)-f(y)|}{|x-y|}$ the Lipschitz norm. Furthermore, in general, for any p,

$$\frac{1}{\kappa}\delta_{1,\kappa}(d\mu_0, d\mu_1) \to d_{\text{TV}}(d\mu_0, d\mu_1) \text{ as } \kappa \to 0.$$

Remark E.3. The transportation problem has indeed some very nice properties that relate to the weak* continuity of the corresponding distance metrics. In particular smoothness with respect to translations and small deformations is an intrinsic property. For this reason it has been used in conjunction with other notions of distance, in a similar fashion as in the current paper, to link density functions of unequal mass. In particular, Benamou and Brenier [4] have introduced a mixed L^2/Wasserstein optimal mapping to link such density functions, while in other relevant literature, Caffarelli and McCann [6] and more recently, Figalli [9] study the transportation of a portion of two unequal masses onto each other.

E.6 Examples

We present two examples that highlight the relevance of the proposed metrics in spectral analysis. The first example compares how different distance measures perform on spectra which contain spectral lines. The second compares how these measures distinguish voiced sounds of different speakers. The distance measures we consider, besides the transportation distance (here $\delta_{1,1}$), are the prediction metric and the Itakura-Saito distance. In both examples the time-series are normalized to have the same variance.

Example E.1. We consider a random process $y_k = \cos(k\theta + \phi) + w_k$ which consists of a sinusoidal component and a zero-mean, unit-variance, white-noise component w_k. Here, θ is taken as a constant, whereas ϕ is assumed random, independent of w_k, and uniformly distributed on $(-\pi, \pi]$. Figure E.1 shows three samples of such a random process for

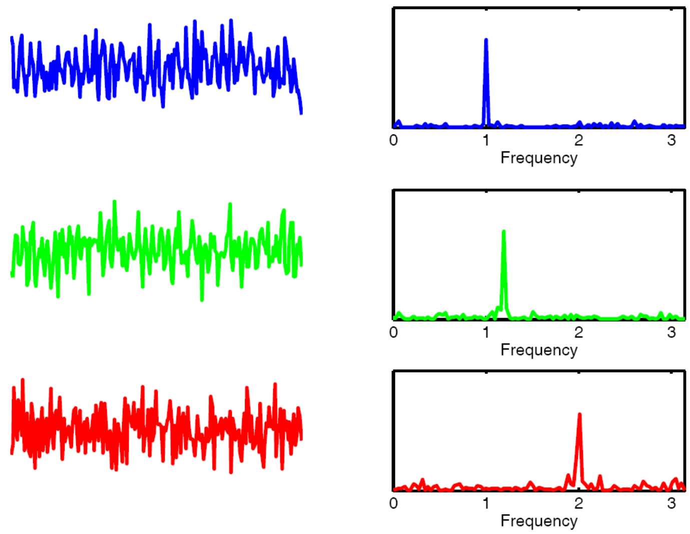

Figure E.1: Stochastic process y_k in time and frequency domain for $\theta = 1, 1.2$, and 2

respective values of $\theta \in \{1,\ 1.2,\ 2\}$, along with their respective power spectra. Based on a set of 500 independent simulations, Table I shows the average distance of the respective power spectra when measured using i) the transport distance, ii) the prediction distance [12], and iii) the Itakura-Saito distance (see e.g., [14]). Comparison of these values reveals that only the transportation-based metric can reliably distinguish between spectral lines.

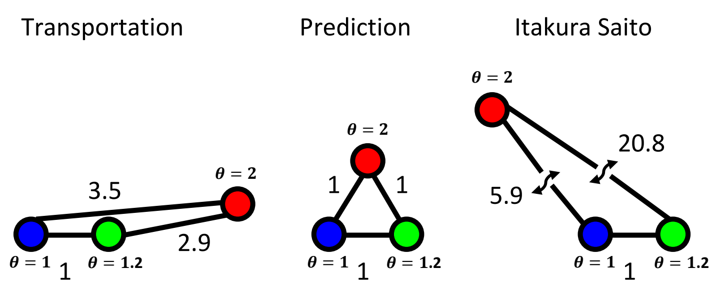

Figure E.2: Relative distances between line spectra

The schematic in Figure E.2 compares the relative distances in these three cases with the smallest value normalized to one. The respective distances for the case of the prediction metric are relatively insensitive to the actual location of the spectral line, as in the limit of a long observation record all three distances ought to be equal to one. Recall that the predic-

Table E.1: Comparison of distance measures on spectral lines + noise

	Distance between line spectra at		
	$\theta_1 = 1$ $\theta_2 = 1.2$	$\theta_2 = 1.2$ $\theta_3 = 2$	$\theta_1 = 1$ $\theta_3 = 2$
Transportation: $\delta_{1,1}$	0.3077	0.8832	1.0690
Predictive: d_{pr}	1.8428	1.8390	1.8517
Itakura Saito: d_{IS}	22.7279	472.6690	134.1707

tion metric does not detect deterministic components. On the other hand the Itakura-Saito distance gives a rather distorted view of reality. In the transportation metric the respective distances are consistent with "physical" location of the spectral lines. Further, the consistency in the ability to discriminate between such spectra is dramatically different in the three cases. Consider the proportion of the simulations for which the distance between the first two spectra ($\theta = 1, 1.2$) is smaller than any of the other distances ($\theta = 1.2, 2$ or $\theta = 1, 2$). For the transport distance in all 500 iterations the distance between the first two power spectra with lines at $\{1,\ 1.2\}$ was smaller than the distance between the other two possibilities. On the other hand, the corresponding percentages for the prediction distance and for the Itakura-Saito distance were 34.4% and 33.8%, respectively. Thus, the transportation correctly identifies the two spectra that are intuitively closest (i.e., having spectral lines closest to each other), whereas the other distance measures succeed about one third of the times (practically a random pick).

To be fair, neither the prediction metric nor the Itakura-Saito distance were designed, or claimed, to have such discrimination capabilities. As the sample size tends to infinity power spectra computed via the periodogram method converge to the true spectrum in weak*, and since the transportation distance is weak*-continuous, transportation distances converge to the true values. On the other hand, this is not the case for either of the other two distances. □

Example E.2. Figure E.3 (left side column) shows time samples corresponding to the phoneme "a" spoken by three individuals, speakers A (Alice), B (Bob), and C (Colin), respectively. Speakers B and C were chosen to be males whereas speaker A was chosen to be a female and, accordingly, the dominant formant of the first speaker has higher pitch than the other two, as seen in the estimated power spectra shown in Figure E.3 (right hand side column). The distance between these three power spectra in the transportation metric, the prediction metric, and the Itakura-Saito distance are compared in Figure E.4 as before. The shortest distance in all three cases is normalized to one.

It is seen that all three distance measures are consistent in that the power spectrum corresponding to Bob is always between that of Alice and Colin. However, the respective distances are highly skewed, especially when it comes to the Itakura-Saito distance. □

In the second example, all three distance measures appear to give qualitatively similar results. However, in general, visual comparison of two power spectra appears to be more

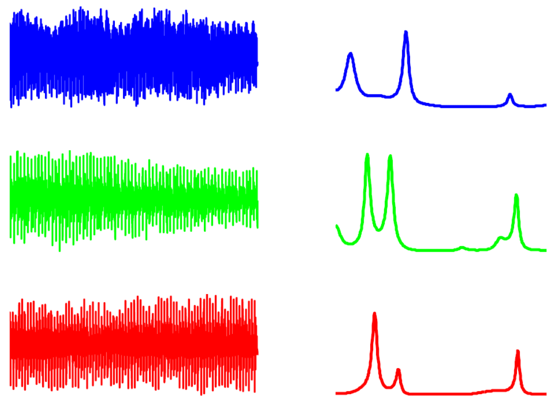

Figure E.3: phoneme “a” for Alice, Bob, and Colin (top to bottom) in time and frequency domain

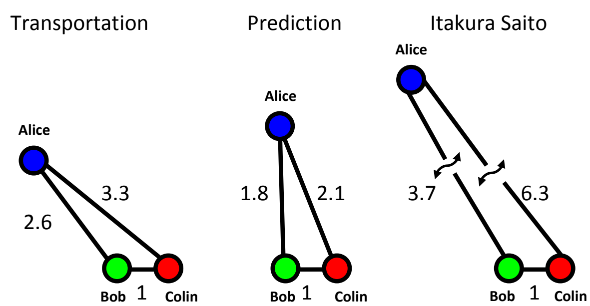

Figure E.4: Relative distances between spectra of the voiced sounds in Figure E.3

easily correlated with respective distances in the transportation metric. This is rather evident in the first example and the reasons are traceable to the physical interpretation of the metric with regard to mass transfer. Moreover, we should point out that in neither example did the Itakura-Saito distance respect the triangular inequality (and of course, it has never been claimed to satisfy this metric property).

E.7 Concluding remarks

Our goal has been to identify natural notions of distance for quantitative spectral analysis. Historically, there has been a variety of options [14, 16, 21, 22] which were used to measure distortion, and which were relied upon based on their perceptive qualities, e.g., see [24, 3]. However, in order to quantify spectral uncertainty, the relevant metrics ought to allow localization of power spectra based on estimated statistics (i.e., being weak*-continuous). At the same time, these metrics ought to share certain natural properties with regard to how noise affects distance between power spectra. In the present paper, we have presented an axiomatic framework that attempts to capture these intuitive notions and we have developed a family of metrics that satisfy the stated requirements.

While there are many possibilities for developing weak*-continuous metrics as suggested, we have chosen to base our approach on the concept of transportation. The reason is that the resulting metrics have certain additional properties which relate to deformations of spectra and smoothness with respect to translation. More specifically, from experience, it appears that geodesics (in e.g., the Wasserstein 2-metric) preserve "lumpiness." A consequence is that when linking power spectra of two similar speech sounds via geodesics of the metric, the corresponding formants often seem to be "matched" and the power between those to transfer in a consistent manner. Such a property appears highly desirable in speech morphing (cf., see [15]). Thus, it will be interesting to expand the set of axioms to include such desirable properties in a more formal way. Such an exercise may in turn narrow down the possible choices of metrics and provide further justification for transportation-based metrics.

Finally, we wish to comment on the need for analogous metrics for comparing multivariable spectra. The ability to localize matricial power spectra is of great significance in system identification, as for instance, in identification based on joint statistics of the input and output processes of a system (cf. [13]).

E.8 Bibliography

[1] B.P. Adhikari and D.D. Joshi. Distance, discrimination et resumé exaustif. *Publ. Inst. Univ. Paris*, 5:57–74, 1956.

[2] S. Amari and H. Nagaoka. *Methods of Information Geometry*. Translations of Mathematical Monographs, AMS, Oxford University Press, 2000.

[3] M. Basseville. Distance measures for signal processing and pattern recognition. *Signal Processing*, 18(4):349–369, 1989.

[4] J.D. Benamou and Y. Brenier. Mixed l^2/wasserstein optimal mapping between prescribed density functions. *J. Optimization Theory and Applications*, 111(2):255–271, November 2001.

[5] A. Bhattacharyya. On a measure of divergence between two statistical populations defined by their probability distributions. *Bull. Calcutta Math. Soc.*, 35:99–109, 1943.

[6] L.A. Caffarelli and R.J. McCann. Free boundaries in optimal transport and monge-ampère obstacle problems. *Annals of Mathematics*, to appear.

[7] N.N. Chentsov. *Statistical Decision Rules and Optimal Inference*. Nauka, Moscow, Providence, 1982. English translation.

[8] N.N. Chentsov. The unfathomable influence of kolmogorov. *The Annals of Statistics*, 18(3):987–998, 1990.

[9] A. Figalli. The optimal partial transport problem. Université de Nice-Sophia Antipolis, preprint, April 2008.

[10] T.T. Georgiou. Distances between power spectral densities. *IEEE Trans. on Signal Processing*, 55(8):3993–4003, August 2007.

[11] T.T. Georgiou. Distances between time-series and their autocorrelation statistics. In A. Chiuso, A. Ferrante, and S. Pinzoni, editors, *Modeling, Estimation and Control*, pages 113–122. Springer-Verlag, Berlin, 2007. Festschrift in Honor of Giorgio Picci.

[12] T.T. Georgiou. An intrinsic metric for power spectral density functions. *IEEE Signal Processing Letters*, 14(8):561–563, August 2007.

[13] T.T. Georgiou, C.R. Shankwitz, and M.C. Smith. Identification of linear systems: A graph point of view. In *American Control Conf*, 1982.

[14] R. Gray, A. Buzo, A. Gray, and Matsuyama. Distortion measures for speech processing. *IEEE Trans. on Acoustics, Speech, and Signal Proc.*, 28(4), August 1980.

[15] X. Jiang, S. Takyar, and T.T. Georgiou. *Metrics and morphing of power spectra*, chapter in Lecture Notes in Control and Information Sciences. Springer Verlag, 2008. to appear.

[16] A. H. Gray Jr. and J. D. Markel. Distance measures for speech processing. *IEEE Trans. Acoustics, Speech, Signal Proc.*, 28(4):380–391, October 1976.

[17] J. Karlsson and T.T. Georgiou. Signal analysis, moment problems & uncertainty measures. In *IEEE Proc. CDC*, pages 5710– 5715, December 2005.

[18] R.E. Kass. The geometry of asymptotic inference. *Statistical Science*, 4(3):188–234, 1989.

[19] L. Knockaert. A class of statistical and spectral distance measures based on bose-einstein statistics. *IEEE Trans. on Signal Processing*, 41(11):3174–3177, November 1993.

[20] S. Kullback. *Information Theory and Statistcs*. Dover, 1997.

[21] J. Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April 1975.

[22] J.D. Markel and A.H. Gray Jr. *Linear prediction of speech*. Springer-Verlag, New York, 1976.

[23] M.S. Pinsker. *Information and Information Stability of Random Variables and Processes*. Holden-Day, 1964.

[24] F. Soong and M.M. Sondhi. A frequency-weighted itakura spectral distortion measure and its application to speech recognition in noise. *IEEE Trans. Acoustics, Speech, Signal Proc.*, 36(1):41–48, January 1988.

[25] C. Villani. *Topics in Optimal Transportation*, volume 58. Graduate studies in Mathematics, AMS, 2003.