ELSEVIER

Journal of Computational and Applied Mathematics 121 (2000) 355–378

JOURNAL OF COMPUTATIONAL AND APPLIED MATHEMATICS

Rational approximation in linear systems and control ☆

A. Bultheel[a, *, 1] , B. De Moor[b, 2]

[a] *Department of Computer Science (NALAG), K.U.Leuven, Celestijnenlaan 200 A, B-3001 Leuven, Belgium*

[b] *Department of Electrical Engineering (ESAT-SISTA), K.U.Leuven, Kardinaal Mercierlaan 94, B-3001 Leuven, Belgium*

Received 7 July 1999; received in revised form 27 January 2000

Abstract

In this paper we want to describe some examples of the active interaction that takes place at the border of rational approximation theory and linear system theory. These examples are mainly taken from the period 1950–1999 and are described only at a skindeep level in the simplest possible (scalar) case. We give comments on generalizations of these problems and how they opened up new ranges of research that after a while lived their own lives. We also describe some open problems and future work that will probably continue for some years after 2000. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Rational approximation; Linear system theory; Model reduction; Identification

1. Introduction

In linear systems, control, and signal processing, rational approximation has always been an important issue and it has given rise to specific problems and insights in approximation theory, it has

☆ This work is partially supported by the Belgian Programme on Interuniversity Poles of Attraction, initiated by the Belgian State, Prime Minister's Office for Science, Technology and Culture. The scientific responsibility rests with the authors.

* Corresponding author. Tel.: +32-16-327540; fax: +32-16-327996.

E-mail addresses: adhemar.bultheel@cs.kuleuven.ac.be (A. Bultheel), bart.demoor@esat.kuleuven.ac.be (B. De Moor)

[1] Partially supported by the Fund for Scientific Research (FWO), project "Orthogonal systems and their applications", grant #G.0278.97.

[2] Research associate of the Fonds Wetenschappetyk Onderzoek Vlaanderen and Associate Professor at the K.U. Leuven. Also supported by grants from the Flemish Government (Research council KUL, GOA-MIPS, FWO-Vlaanderen (grants, research projects and communities), AWI (Bil. Int. Coll.), IWT (Eureka SINOPSYS, Eureka Impact, STWW)), the Belgian Federal Government (IUAP-IV/02 and IUAP-IV/24, Program Dur. Dev. (Mobility)), the European Commission (TMR Networks (Alapedes, Niconet), Science (ERNSI)) and several industrial grants (Electrabel, Laborelec, Verhaert).

0377-0427/00/$ - see front matter
PII: S0377-0427(00)00339-3

revived forgotten methods and initiated new directions of research. It is the intention of this paper to illustrate some of these innovating ideas that were born from this interaction of system theory, linear algebra and approximation theory and formulate some open problems or aspects that are still under development.

First we recall some mathematical notation and concepts. Next, we shall narrow the focus of our system theoretic problems by introducing some concepts from system theory and selecting the subjects we shall discuss from those we shall not consider in this paper.

By $\mathbb{Z}$ we denote the set of integers, by $\mathbb{N}$ the positive integers, and $\ell_p = \ell_p(\mathbb{Z})$, $(1 \leqslant p < \infty)$ is the Banach space of complex sequences $x = (x_k)$ with $\|x\|_p := [\sum_{k\in\mathbb{Z}} |x_k|^p]^{1/p} < \infty$ while ℓ_∞ is the space for which $\|x\|_\infty := \sup_{k\in\mathbb{Z}} |x_k| < \infty$. The set of real and complex numbers is $\mathbb{R}$ and $\mathbb{C}$, respectively, and the notation $\mathbb{T}$, $\mathbb{D}$ and $\mathbb{E}$ are reserved for the unit circle, its interior and its exterior: $\mathbb{T} := \{z \in \mathbb{C}: |z| = 1\}$, $\mathbb{D} := \{z \in \mathbb{C}: |z| < 1\}$, $\mathbb{E} := \{z \in \mathbb{C}: |z| > 1\}$. The spaces $L_p = L_p(\mathbb{T})$, $(1 \leqslant p < \infty)$ are defined as the function spaces for which $\|F\|_p := [(1/2\pi)\int_{-\pi}^{\pi} |F(e^{i\omega})|^p \,d\omega]^{1/p} < \infty$, and $\|F\|_\infty := \sup_{t\in\mathbb{T}} |F(t)|$. L_2 is a Hilbert space with inner product $\langle F, G\rangle = (1/2\pi)\int_{-\pi}^{\pi} F(e^{i\omega})\overline{G(e^{i\omega})}\,d\omega$. The $\mathscr{Z}$-transform of a sequence $a = (a_k)_{k\in\mathbb{Z}}$ is $A(z) = \mathscr{Z}(a) = \sum_{k\in\mathbb{Z}} a_k z^{-k}$. We will use the convention that $\mathscr{Z}$-transforms are indicated by capital letters: $\mathscr{Z}(a) = A(z)$. Note that we use here the system engineering convention that the $\mathscr{Z}$-transform is defined as above, while the mathematical convention is that z is replaced by z^{-1}. We shall define the Fourier transform correspondingly. Thus $\mathscr{F}(a) = A(e^{i\omega}) = \sum_{k\in\mathbb{Z}} a_k e^{-ik\omega}$. The Fourier transform is an isometric isomorphism between ℓ_2 and L_2. Although an integrable function $F \in L_1$ has a Fourier series $\sum_{k\in\mathbb{Z}} f_k e^{-ik\omega}$ with $f_k = (1/2\pi)\int_{-\pi}^{\pi} F(e^{i\omega})e^{ik\omega}\,d\omega$, the partial sums need not converge in norm or pointwise. The Cesàro sums are summation techniques, for example the Fejér sums that take the average over the first n partial sums, and these have better convergence properties than the ordinary partial sums. The Hardy space $H_p = H_p(\mathbb{D})$ $(1 \leqslant p \leqslant \infty)$ is the subspace of L_p of the functions F whose *positive* Fourier coefficients f_k with $k \geqslant 1$ vanish. Because the series $\sum_{k=0}^{\infty} f_{-k} z^k$ converge for $z \in \mathbb{D}$, these functions have an analytic extension in $\mathbb{D}$. Similar to Hardy spaces, one can define a closed subset of the continuous functions on $\mathbb{T}$ namely $\mathscr{A} = \mathscr{A}(\mathbb{D}) = \{F \in C(\mathbb{T}): f_{-n} = \int_0^{2\pi} F(e^{i\omega})e^{in\omega}\,d\omega = 0,\ n = 1, 2, \ldots\}$ which is the *disk algebra*. Again such functions can be extended analytically to $\mathbb{D}$. The closure of the polynomials in L_p give H_p for $1 \leqslant p < \infty$ but it gives $\mathscr{A}$ for $p = \infty$. The inclusion $\mathscr{A} \subset H_\infty$ is proper.

To describe the system theoretic topics of this paper, let us now introduce some terminology from system theory. Mathematically, a *linear system* is a linear operator transforming an input into an output. Depending on the spaces where the input and the output live, there are several possibilities. A system has an input signal u_t say, and an output signal y_t. The variable t stands for "time". If t ranges over a continuous set (like $\mathbb{R}$) we have a *continuous time* system, otherwise (e.g., $t \in \mathbb{Z}$) we have a *digital* or *discrete time* system. Here we shall restrict ourselves to *discrete time* systems.

The system can be *single input–single output* (SISO) if the input and output are scalar. Otherwise, if they are in higher dimensional vector spaces, it is a *multi-input–multi-output* system. We restrict ourselves mainly to the simple SISO case.

There are two settings for these systems: the I/O signals can be stochastic or they can be purely deterministic, or even a mixture, like a deterministic signal with stochastic noise. In this paper, we look mainly at the purely deterministic case.

So we have excluded a lot of possible problems, but the choice that remains is still overwhelming. For the areas that are not treated, a specialist would know other problems analogous to those discussed here. The nonspecialist will find enough material to start with.

In the simplest possible form, a system can be described as a convolution: $y = T_h(u) = h * u$, thus $y_k = \sum_i h_{k-i} u_i$, $k \in \mathbb{Z}$, where $u = (u_k)$, $y = (y_k)$ are the input and output signals, and $h = (h_k)$ is called the *impulse response* of the system. The h_k are also known as *Markov parameters*. Taking $\mathscr{Z}$-transforms, we can write the I/O relation as $Y(z) = H(z)U(z)$. If the series $H(z)$ converges, then $H(z)$ represents a function that is called the *transfer function* of the system.

Unless otherwise implied, we shall assume that we are dealing with *causal* systems, which means that the impulse response $h_k = 0$ for $k < 0$. If a sequence (s_k) has finite *energy* $\sum_k |s_k|^2 < \infty$, then it means that $s \in \ell_2$. A system is called *stable* if the I/O operator $T_h : u \mapsto y = T_h(u)$ is bounded. If $u, y \in \ell_2$ then the system is *stable* if $h \in \ell_\infty$, in other words, according to our convention, a stable causal ℓ_2 system will have a transfer function H such that $H(1/z)$ is a bounded analytic function in the open unit disk: $H \in H_\infty(\mathbb{D})$.

The I/O signals (sequences) live in the *time domain* while the $\mathscr{Z}$-transforms live in the $\mathscr{Z}$ domain or frequency domain. Indeed, if $z = e^{i\omega}$ in the $\mathscr{Z}$-transform, then $H(z)$ is just the Fourier transform of $h = (h_k)$. The function $H(e^{i\omega})$ is called the frequency response and $|H(e^{i\omega})|^2$ is the *power spectrum*.

In many cases, it is very convenient to write a state-space formulation for the system. This means to define an intermediate vector x_k called the *state* such that

$$x_{k+1} = Ax_k + Bu_k, \quad u_k, y_k \in \mathbb{C},\ x_k, x_{k+1} \in \mathbb{C}^{d\times 1},$$

$$y_k = Cx_k + Du_k, \quad A \in \mathbb{C}^{d\times d},\ B \in \mathbb{C}^{d\times 1},\ C \in \mathbb{C}^{1\times d},\ D \in \mathbb{C},$$

where we shall assume that the dimension d of the *state space* is finite. The relation between state-space, time-domain, and frequency-domain formulations are (under appropriate conditions) given by

$$H(z) = C(zI - A)^{-1}B + D = \sum_k h_k z^{-k},$$

$$h_0 = D, \quad h_k = CA^{k-1}B, \quad k = 1, 2, \ldots .$$

The state space is a space intermediate between input and output space. The previous equations for x_{k+1}, and y_{k+1} show one way in which the input is mapped to the state and the state is mapped to the output. In control problems it is of great importance to know how much of the state space can be reached from the input side and how much from the state space can be read off from the output. This is characterized to some extent by the controllability and observability Gramians, respectively. The controllability matrix $\boldsymbol{C}$ is the array whose kth column is $A^{k-1}B$, $k = 1, 2, \ldots$ and the observability matrix $\boldsymbol{O}$ is the array whose kth row is CA^{k-1}, $k = 1, 2, \ldots$. The corresponding Gramians are $P = \boldsymbol{C}\boldsymbol{C}^*$ and $Q = \boldsymbol{O}^*\boldsymbol{O}$. They solve the Lyapunov equations $APA^* - P = -BB^*$ and $A^*QA - Q = -C^*C$, respectively. (The superscript * means conjugate transpose.) Another important tool is the Hankel operator Γ_H of the system (see Section 3.2) which maps the past input onto the future output. With respect to the standard basis, it has a Hankel matrix representation $[h_{i+k-1}]_{i,j=1}^{\infty}$, which we can also write as $\boldsymbol{O}\boldsymbol{C}$.

The state-space formulation is very important for numerical computations and it is the most elegant way of dealing with generalizations like block Hankel matrices or MIMO systems or time-varying systems. State-space methods involve basically linear algebra techniques. Many of the most practical algorithms are based on state-space descriptions and linear algebra techniques. However, conceptually it is sometimes easier to work in function spaces. So, in order not to develop the ideas twice, we have made a choice in the different sections of this paper: sometimes we give a state-space description, sometimes the linear algebra aspects are deliberately neglected.

The general problem that is addressed here is:

Problem 1.1. (General problem) *Given some data about the system, find* (*a rational*) *approximation of the system.*

Of course, this problem has several "names". For example, in *model reduction* the aim is to find a simple (i.e., a low degree) approximation of a high-order system. In *realization* problems the objective is to find an explicit, implementable form of the system. Most often this is meant to be a state-space realization. *Identification* of the system means again that the system is pinned down in some way, and this is usually meant to be in the time domain or in the frequency domain. For instance, in the time domain this means that a sequence of (possibly perturbed) inputs u_k and/or outputs y_k is known. *Prediction* theory (or *forecasting*) is vaguely described as trying to identify or model or approximate a system for which we know only the output (like a seismic or a speech signal). In *control* theory one wants a system to generate a certain output, and if the actual output deviates from the desired one, then the input or the system itself is modified (controlled). To do this in an effective way, it is of course necessary to have at least an approximate model of the system, and thus the previous problems reappear.

The data that are given for the system can also have many different forms. Sometimes data are given for the behaviour of $H(z)$ for z near ∞ (e.g., the first few Markov parameters $h_k = CA^{k-1}B$ which are the coefficients in the expansion of $H(z)$ at $z = \infty$, which describe best the steady state behavior of the system, i.e., the behavior of the system for large t), or at the origin (e.g., the so-called time moments, or equivalently the coefficients of $H(z)$ expanded at $z = 0$ which describe best the transient behavior of the system, i.e., the behavior of the system for small t), or on the unit circle (e.g., the power spectrum $|H(z)|^2$ for $|z| = 1$). These data could be assumed exact, but they are in all practical situations contaminated by measurement error or model errors (for example nonlinear effects, while the model is linear, or an underestimation of the model order). Not only the data and the desired result but also the *approximation criterions* can be formulated in the time or frequency domain. That may be some kind of norm (like L_p norms) or some other criterion like interpolation with minimal degree or minimal norm, etc.

We want to give some samples from the last 50 years that we consider as stepping stones in the fertile interaction between system theory and rational approximation in the complex plane. We emphasize that it is never our ambition to be complete, and we do not want to deprecate any work we do not mention. We give some examples that are obviously colored by our personal interest. It should be sufficient though to illustrate our main point, namely the rich soil between system theory, rational approximation, linear algebra, numerical analysis, and operator theory that has been an incubation place for many new ideas and methods and will continue to be so in many years to come.

2. Realization theory

One of the main achievements in system theory of the 1960s was the solution of the so-called *minimal partial realization* problem.

Problem 2.1. (Minimal partial realization) *Given the expansion* $H(z)=h_0+h_1z^{-1}+\cdots+h_Nz^{-N}+\cdots$ *for* $z\to\infty$. *Find a rational function* $\hat{H}$ *of minimal degree satisfying* $H(z)-\hat{H}(z)=\mathrm{O}(z^{-(N+1)})$ *as* $z\to\infty$.

Writing $\hat{H}(z)=P_n(z)/Q_n(z)$ with P_n and Q_n polynomials, one can solve the linearized problem

$$H(z)Q_n(z)-P_n(z)=\mathrm{O}(z^{-(N+1)}),\quad z\to\infty$$

instead of the proper problem. If this is written down coefficientwise in descending order, we see that the first $n+1$ equations give P_n in terms of Q_n and $h_0,\ldots,h_n$. The most obvious choice is $N=n$, so that generically the next n equations define the coefficients of Q_n, up to a multiplicative factor, as the solution of a homogeneous Hankel system of linear equations. If this Hankel matrix is not of full rank there can be more solutions. Then the ratio $\hat{H}$ is still unique, but it may or may not be a solution of the proper problem.

This is equivalent with the construction of a diagonal Padé approximation at infinity [2]. In Padé approximation one considers rational approximants of general type (m,n) [that is (numerator, denominator)-degree] whose expansion matches a given series in its first $m+n+1$ terms. The (m,n)-approximants are arranged in a table called Padé table.

The problem can be solved recursively, i.e., it is computationally cheap to update a solution for a given order N to a solution for order $N+1$. This kind of algorithms corresponds to fast algorithms for the solution of Hankel matrices that were recognized as variants of the Euclidean algorithm [14]. These algorithms solve a Hankel system by solving all the systems corresponding to the leading principal submatrices of the Hankel matrix. In this sequence there may appear singular submatrices according to a certain "pattern" [29]. This phenomenon is well known in Padé theory and corresponds to singular blocks in the Padé table. Such singular blocks are always square and that explains the "pattern".

These algorithms are fast because they exploit the structure of the Hankel matrix. However, the linear algebra operations do not use orthogonal matrices and to maintain the structure of the matrix, pivoting is not allowed. Therefore, these algorithms potentially suffer from numerical instabilities. The leading principal submatrices may be nearly singular, which causes large rounding errors. This problem has initiated an intensive research in the 1990s about look-ahead techniques for such structured linear algebra problems. These algorithms test if some quantity is below a certain threshold, and in that case the matrix is considered as being singular, and an update in the recursion is postponed until the quantity raises above the threshold [25].

The coefficients that are computed by such an algorithm correspond to recurrence relations for formal orthogonal polynomials [6] and to the numerators and denominators in a continued fraction expansion of the transfer function, which are known in realization theory as *Cauer* fractions.

Note, however, that this technique does not guarantee stability of the approximant and that can be considered as a major drawback of this technique. The coefficients of the recurrence relation for the orthogonal polynomials, or equivalently of the continued fractions, can be used in a stability test.

This is the so-called Routh algorithm: a method to test whether a certain polynomial is stable (in the sense of continuous time systems, i.e., having zeros in the left half-plane) or not. The parameters in the Routh algorithm can be restricted, such that the rational approximant that results is guaranteed to be stable. Another possibility is to start from a high-order denominator polynomial of a stable system. The Routh algorithm computes the Routh parameters and by truncation, one obtains a low degree denominator polynomial for the approximation. The free coefficients in the numerator are used to match as many Markov parameters as possible. This is an example of a Padé-type approximant: the denominator is given and the numerator is determined in Padé sense.

More on the topic of minimal partial realization and on Padé and multivariate Padé approximation can be found in the extended contributions in this volume by De Schutter, Wuytack, and Guillaume, respectively.

3. Model reduction techniques

In the model reduction problem one wants to approximate a linear system by one of lower McMillan degree. The methods used here are diverse and rely on different mathematical techniques, on different presentations of the system and on different objectives that are optimized. We give some examples.

3.1. Balanced truncation and projection of dynamics

The idea is to perform a similarity transformation on the state space, so that the observability and controllability are "in balance". This means that the controllability Gramian P and the observability Gramian Q are equal and diagonal (see [40]). In this balanced realization, a simple truncation (keeping the most important eigenvalues) does the actual model reduction. The balanced realization is obtained from an eigenvalue decomposition of the product $PQ = T\Lambda T^{-1}$. Assume that the d eigenvalues in Λ are ordered in decreasing order: $\Lambda = \mathrm{diag}(\Lambda_1, \Lambda_2)$ with Λ_1 containing the $n \leqslant d$ largest eigenvalues. After the similarity transformation $(A, B, C, D) \to (TAT^{-1}, TB, CT^{-1}, D)$, we then keep only the the first n rows and/or columns to isolate the reduced system. The reduced system is stable if the original system is, but the reduced system is not balanced in the discrete case (it is balanced for continuous time systems though). We note that the eigenvalues of PQ are related to the Hankel singular values. These are the singular values of the Hankel matrix of the system $\Gamma_H = \boldsymbol{OC}$ which maps past input into future output (see Section 3.2). Indeed one has

$$\lambda_i(PQ) = \lambda_i(\boldsymbol{CC}^*\boldsymbol{O}^*\boldsymbol{O}) = \lambda_i(\boldsymbol{C}^*\boldsymbol{O}^*\boldsymbol{OC}) = \sigma_i^2(\Gamma_H).$$

Thus, this technique of model reduction typically throws away the smaller Hankel singular values and keeps the most important ones.

There are also several extensions. Actually, whenever one has a theory in which two matrix equations (Lyapunov, Riccati) result in two positive-definite solutions P and Q, one can define a balancing transformation T, a so-called contragredient transformation, that transforms P and Q so that they are equal and diagonal. Examples are stochastic balancing (in which case the solutions to the so-called forward and backward Riccati equations of a stochastic system are used to obtain the balancing transformation, LQG-balancing (where one starts from the solutions of the Kalman filter

and LQR Riccati equations, balancing on the two solutions of the H_∞ Riccati filtering and control equations, relative error balancing (see, e.g., [44,47] for details and references).

Balanced model reduction can be considered as one particular case of the so-called *projection of dynamics*. Hereto define the projection $\Pi = RL$, where $R, L^* \in \mathbb{C}^{d\times n}$ and $LR = I_n$. Π is the projection onto the range space of R along the null space of L. Now, the idea of projection of dynamics is that at each time instant, the state $x(t)$ is approximated by $\Pi x(t) = RLx(t)$ so that

$$RLx_{k+1} \approx ARLx_k + Bu_k, \quad y_k \approx CRLx_k + Du_k,$$

or

$$z_{k+1} = (LAR)z_k + (LB)u_k, \quad y_k = (CR)z_k + Du_k$$

with $z_k := Lx_k$.

Since $LR = I_n$, it is always possible to find a similarity transformation T such that L consists of the first n rows of T and R of the first n columns of T^{-1}. Hence the approximating system (LAR, LB, CR, D) is obtained by first transforming the model by choosing a particular basis transformation matrix T, giving the realization $(TAT^{-1}, TB, CT^{-1}, D)$, and then truncating the transformed state-space model by restriction to the first n rows and/or columns. The oldest methods that can be interpreted in terms of *projection of dynamics* are *modal decomposition and reduction methods*, in which case the similarity transformation T diagonalizes the matrix A. The eigenvalues of the matrix A, which are in case of a minimal realization also the poles of the system, are thus revealed. This allows us to easily choose the poles of the reduced order system as a subset of the poles of the original system, or sometimes poles of the reduced system are simply fixed in advance. This is similar to the Padé-type approximants mentioned at the end of Section 2. The advantage of these method is their simplicity, typically only requiring the solution of one or more least-squares problems.

3.2. Hankel norm approximation

The Hankel norm approximation problem was inspired by a paper of Adamjan, Arov and Krein that appeared in 1971 [1] and it is therefore also known as AAK approximation. It was worked out in the context of system theory in the late 1970s and early 1980s.

If the system is stable and causal, then the I/O operator T_h defined in Section 1 will map the past onto the future. If we assume that the input and output have finite energy, then T_h is a Hankel operator in ℓ_2. In the $\mathscr{Z}$ or Fourier domain, this corresponds to a Hankel operator Γ_H mapping $H_2 = H_2(\mathbb{D})$ (the Hardy space of the unit disk) to its orthogonal complement $H_2^\perp$. It is defined by $\Gamma_H = P_{H_2^\perp} M_H$, where $H(z) = \sum_{k=1}^{\infty} h_k z^{-k}$ for $z \in \mathbb{T}$ is the Fourier transform of the impulse response h, and M_H is the operator representing the multiplication with H, and $P_{H_2^\perp}$ is the orthogonal projection onto $H_2^\perp$. If the system is stable, then Γ_H is a bounded operator and hence $H \in L_\infty$. The function $H \in L_\infty$ is called the *symbol* of the Hankel operator Γ_H. Note that only the Fourier coefficients h_k with $k > 0$ are relevant to define the operator. Given a Hankel operator, then its symbol is only determined up to an additive arbitrary H_∞ function.

The representation of Γ_H with respect to the standard bases $\{z^k\}$ is a Hankel matrix whose entries are the Markov parameters.

Problem 3.1. (Hankel norm/AAK approximation) *Given Γ_H, find an approximation $\Gamma_{\hat{H}}$ such that we solve one of the following two problems*:

- *minimum norm problem*: $\|\Gamma_{H-\hat{H}}\|_2$ *is minimal with* $\operatorname{rank}\Gamma_{\hat{H}} \leqslant n$, *or the*
- *minimum rank problem*: $\operatorname{rank}\Gamma_{\hat{H}}$ *is minimal with* $\|\Gamma_{H-\hat{H}}\|_2 \leqslant \varepsilon$ *for given* ε.

Of course, the minimal norm problem has a solution for general matrices that is given by a singular value decomposition (SVD), truncated after the first (i.e., the largest) m singular values. Thus if $\Gamma = \sum_k \sigma_k \langle \cdot, v_k \rangle w_k$ is the SVD of Γ with singular values $\sigma_1 \geqslant \sigma_2 \geqslant \cdots$, and corresponding Schmidt pairs (v_k, w_k), then the best rank m approximant is given by keeping only the first n terms in this sum. The remarkable fact here is that for a Hankel matrix Γ_H, where this truncated approximation is in general not a Hankel matrix, we can find a Hankel approximant $\Gamma_{\hat{H}}$ that approximates equally well. Thus $\inf\{\|\Gamma_H - \Gamma_{\hat{H}}\|_2 \colon \operatorname{rank}\Gamma_{\hat{H}} \leqslant n\} = \sigma_{n+1}$.

The rank of any Hankel operator Γ_H is related to the degree of its symbol H by Kronecker's theorem (1890) which says that $\operatorname{rank}\Gamma_H \leqslant n$ iff $H \in \mathscr{R}_n + H_\infty$ where $\mathscr{R}_n$ is the subset of all rational L_∞ functions with at most n poles inside the open unit disk $\mathbb{D}$.

It is clear that if $\Gamma_{\hat{H}}$ approximates Γ_H, then $\hat{H}$ should approximate H in some sense. However, the symbol of a Hankel operator can be in L_∞ in general and therefore we have to say that the "Hankel norm" defined as $\|H\|_\Gamma = \|\Gamma_H\|_2$ is not really a norm for $H(z)$, unless we know that H has no component in H_∞. This Hankel norm is however closely related to the L_∞ norm. A theorem by Nehari (1957) says, for example, that $\|H\|_\Gamma = \|\Gamma_H\|_2 = \inf\{\|H - F\|_\infty \colon F \in H_\infty\} = \operatorname{dist}(H, H_\infty)$. So we arrive at the AAK theorem in its simplest form, which solves the minimal norm problem and gives an approximation result for the symbols of Hankel matrices.

Theorem 3.1. (Adamjan et al. [1]) *Let Γ_H be a compact Hankel operator with Schmidt pairs (v_k, w_k) for the singular values σ_k. Then, with the notation introduced above*

$$\inf\{\|\Gamma_H - \Gamma_{\hat{H}}\|_2 \colon \operatorname{rank}\Gamma_{\hat{H}} \leqslant n\} = \inf\{\|H - \hat{H}\|_\infty \colon \hat{H} \in \mathscr{R}_n + H_\infty\} = \sigma_{n+1}.$$

Let us introduce the $\mathscr{Z}$ transforms $V_k = \mathscr{Z}(v_k)$ and $W_k = \mathscr{Z}(w_k)$. Then if $\sigma_n > \sigma_{n+1}$, there is a unique solution $\hat{H}$ that is defined by

$$H - \hat{H} = \frac{\sigma_{n+1} W_{n+1}}{V_{n+1}} = \frac{\Gamma_H V_{n+1}}{V_{n+1}}.$$

The error function $E = H - \hat{H}$ satisfies $|E| = \sigma_{n+1}$ a.e. on $\mathbb{T}$.

Note that the solution gives a best approximation $\hat{H}$ that is in L_∞, even if the given H has no H_∞ part. So, to have a causal stable approximant, one should get rid of the H_∞ part of $\hat{H}$. Since $\hat{H}$ is rational, this could, in principle, be done by partial fraction decomposition, although this is not the most advisable way to be followed for numerical computations.

The trouble is that for the solution of the AAK problem, one needs to solve an SVD problem for an infinite Hankel matrix. However, when the rank of that matrix is finite, then the computations can be done on a finite number of data like, for example, the state-space description of a matrix. The most elegant solution of the problem came therefore from the state-space approach by Glover [26], which is a benchmark paper in this theory. The operation of balancing the state-space representation of the linear system, as briefly explained previously, is crucial in the state-space approach by Glover.

In this paper it can also be found that generically (when all the Hankel singular values σ_i are distinct) one has for a system of degree d that is approximated by a system of degree n that $\|H-\hat{H}\|_\infty \leqslant 2\sum_{k=n+1}^{d}\sigma_i(\Gamma_H)$.

3.3. H_2 -model reduction

The Hankel norm approximant discussed in the previous section models an approximating system that minimized the H_2-deviation of the future outputs, given that the original and reduced system had the same past inputs. The H_2-norm of the system is the H_2-norm of the transfer function $\|H\|_2$ which equals the ℓ_2-norm of the impulse response $(\sum_k |h_k|^2)^{1/2} = \|h\|_2$. For a finite-dimensional state space, one can derive that this norm can also be expressed in terms of state-space quantities as $\|H\|_2^2 = B^*QB + D^*D = CPC^* + DD^*$. The physical interpretation of the H_2-norm is that its square is the expected value of the power in the output signal, when the input is white Gaussian zero mean noise with unit variance, or in a deterministic model, a unit impulse. Approximation in the H_2-norm means finding a least-squares approximation of the impulse response. It is also known that if $\hat{H}_n$ is an optimal H_2 approximant with a pole p, then $(\mathrm{d}^m/\mathrm{d}z^m)(H(z)-\hat{H}_n(z))_{z=p^{-1}} = 0$, $m = 0, 1$. This necessary condition can and has been be exploited in algorithms. Compare with Section 4 on linear prediction where interpolation problems also satisfy least-squares optimality conditions. We shall stick here to the straightforward characterization, and so our problem is

Problem 3.2. (h_2 norm approximation) *Given a system with transfer function H or with state space (A,B,C,D). Find a system with transfer function $\hat{H}$ or state space $(\hat{A},\hat{B},\hat{C},\hat{D})$ such that $\|H-\hat{H}\|_2$ is minimal where the approximating system can have degree n at most.*

Finding $\hat{H}_n = \arg\min\|H-H_n\|_2$ where H_n ranges over all systems of degree n at most is a classical least-squares problem that is typically solved using an orthogonal basis and orthogonal projection techniques. It can be proved that the solution will generically have degree n since the minimum strictly decreases as n increases. This result is valid even for local minimizers. The H_2-norm squared is clearly differentiable everywhere and therefore necessary conditions for minimality may easily be derived by calculating the gradient of the square of the H_2-norm with respect to the state-space matrices of the candidate approximant $(\hat{A},\hat{B},\hat{C},\hat{D})$. It can also be shown that the state-space model of an H_2-optimal approximant can be characterized in terms of *projection of dynamics*. This characterization is useful in deriving gradient-based optimization algorithms (none of which can, of course, guarantee that a global minimum will be found) or in deriving *homotopy*-based optimization methods. For an excellent survey and many references we refer to [44].

A quite interesting derivation of properties, characterization of optimality, and even algorithms goes via linear algebra. Here we reconsider the Markov parameters $h_k = CA^{k-1}B$, $k = 1, 2, \ldots$. Let $p(\lambda)$ be the characteristic polynomial[3] of A, i.e., $p(\lambda) = \det(\lambda I - A) = \lambda^n + \alpha_1\lambda^{n-1} + \cdots + \alpha_{n-1}\lambda + \alpha_n$. Then, it follows from the Cayley–Hamilton theorem that the h_k satisfy a nth-order recurrence relation, and thus the $p \times q$ Hankel matrix $\mathscr{H}_{p,q} = [h_{i+j-1}]_{i=1,\ldots,p}^{j=1,\ldots,q}$ is rank deficient, when both $p > n$ and $q > n$.

[3] We do not make the distinction here between the characteristic and the minimal polynomial. 'Generically', they coincide. But if not, the presentation here can be refined.

Now, let $h_1, h_2, h_3, \ldots, h_N$ be N given data. We would like to model these datapoints by the impulse response of an nth-order linear system, where n is (a typically small integer) specified by the user. Using the rank deficiency result of the Hankel matrix, we can formulate this least-squares approximation problem as $\min \sum_{k=1}^{N} (h_k - \hat{h}_k)^2$, where the minimum is taken over all $\hat{h}_1, \ldots, \hat{h}_N$, subject to the condition that $\hat{\Gamma}$, the Hankel matrix of size $(N-n) \times (n+1)$ with symbol $\hat{H}$, is rank deficient, i.e., $\hat{\Gamma} v = 0$, for some v, with $\|v\|_2 = 1$. When $N \to \infty$, this problem becomes the optimal H_2-approximation problem we considered above. Let us however first treat the case where $N < \infty$. Then the problem is a so-called *structured total least-squares* problem, the structure being that $\hat{\Gamma}$ is required to be a Hankel matrix. The solution to this follows from the smallest singular value and corresponding singular vector of a nonlinear generalized SVD of the Hankel matrix $H \in \mathbb{R}^{(N-n)\times(n+1)}$ with the data:

$$Hv = D_v u\tau, \quad u^* D_v u = 1,$$

$$H^* u = D_u v\tau, \quad v^* D_u v = 1$$

and $v^* v = 1$, which is called the Riemannian SVD [19]. Specifically for the Hankel case, we refer to [18,20]. In these expressions, D_u and D_v are symmetric weighting matrices of appropriate dimensions and with a particular structure, the elements of which are quadratic functions of the elements of u, resp. v. For all values of u and v they are positive definite. More specifically, in this case D_u and D_v are symmetric, positive-definite banded Toeplitz matrices, the elements of which are quadratic in the components of u and v. The least-squares approximant $\hat{H}$ of McMillan degree n now follows from the 'smallest' singular triplet (u, τ, v) that satisfies the Riemannian SVD equations, as $\hat{h} = h - (u\tau) * v$, where a $*$ denotes the convolution of the vector sequences. Heuristic algorithms, remeniscent of and inspired by the power method to find the smallest singular value of a matrix, are described in [44,17]. For a translation to the $\mathscr{Z}$-domain of the H_2 problem and the Riemannian SVD problem as $N \to \infty$ we also refer to [44].

3.4. The Hilbert–Schmidt–Hankel norm

The Hilbert–Schmidt–Hankel norm is yet another performance criterion. It is defined as the Frobenius norm of its Hankel operator: $\|H\|_{\mathrm{HSH}}^2 = \|\Gamma_H\|_F^2$. It immediately follows from results in the previous sections that for a system of degree d

$$\|H\|_{\mathrm{HSH}}^2 = \sum_{i=1}^{n} \sigma_i^2(\Gamma_H) = \mathrm{Trace}(PQ) = \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} h_i \bar{h}_j = \sum_{i=1}^{\infty} i|h_i|^2.$$

The last equality implies the interpretation of the HSH-norm of a system as a time-weighted H_2-norm: It can be considered as the energy storage capacity, i.e., the expected energy stored in the system at a certain time, when the system has been driven by white zero mean unit variance Gaussian noise up to that time [44, p. 28].

A relatively little known interpretation of the HSH-norm is that it is the area enclosed by the oriented Nyquist plot of the linear system in the complex plane (see [32] for an elaboration and a proof). Hence, when doing model reduction in the HSH-norm, one tries to minimize the area between the Nyquist plot of the given system $H(z)$ and its HSH-optimal approximant of lower McMillan degree.

From what has been discussed about model reduction so far, it should be clear that the essence of model reduction in the HSH norm is again a problem of finding a least-squares rank deficient approximation of a Hankel matrix, that is still a Hankel matrix. It thus should come as no surprise that this problem is again a so-called structured total least-squares problem. For further details, see [44].

Finally, we mention that for SISO systems in discrete time, we have $\|H\|_2 \leqslant \|H\|_\infty$, but there is no upper bound for the H_∞-norm in terms of the H_2-norm. This means that these two norms are not equivalent. However the HSH-norm, the Hankel norm and the H_∞-norm are equivalent:

$$\|H\|_\Gamma \leqslant \|H\|_{\mathrm{HSH}} \leqslant \sqrt{n}\|H\|_\Gamma \quad \text{and} \quad \frac{1}{2n}\|H\|_\infty \leqslant \|H\|_{\mathrm{HSH}} \leqslant \sqrt{n}\|H\|_\infty,$$

where n is the degree of H (see, e.g., [44]).

4. Linear prediction

The aspect of prediction theory is even older and dates back to the theory of Wiener [48], Wiener-Masani [49], Grenander and Szegő [30], and it was revived in the 1980s by Dewilde and Dym [21]. The problem can be described as follows [12]. Suppose that for an arbitrary n we observe a stationary signal $\{h_k\}_{k=-\infty}^{n-1}$ up to time moment $n-1$ and we want to predict the next value h_n. In other words, we should find a model for the signal h. Assume that h_n is predicted as $\hat{h}_n = -\sum_{k=1}^{\infty} a_k h_{n-k}$, then the prediction error $e_n = h_n - \hat{h}_n$ should be minimized in ℓ_2 sense. Taking $\mathscr{Z}$ transforms, we get $E(z) = A(z)H(z)$, where $A(z) = 1 + a_1 z^{-1} + a_2 z^{-1} + \cdots$. Minimizing the L_2-norm $\|E\| = \|AH\| =: \|A\|_{|H|^2}$ leads to a weighted least-squares problem formulated and solved by Szegő. To give an easy formulation, we introduce $F = A_*$ where the substar is used to denote the parahermitian conjugate $A_*(z) = \overline{A(1/\bar{z})}$. Note that $F(0) = 1$.

Problem 4.1. (Szegő) *Given some positive measure μ on $\mathbb{T}$, find a function $F \in H_2^\mu$ that solves* $\inf\{\|F\|_\mu^2 : F \in H_2^\mu, F(0) = 1\}$.

The norm is taken in the Hilbert space L_2^μ with inner product

$$\langle f, g\rangle_\mu = \int_{\mathbb{T}} f(t)\overline{g(t)}\,\mathrm{d}\mu(t)$$

and H_2^μ is the subspace of functions analytic in $\mathbb{D}$. In the prediction problem, the measure μ is called spectral measure since the weight is the power spectrum: $\mathrm{d}\mu(\mathrm{e}^{\mathrm{i}\omega}) = (|H(\mathrm{e}^{\mathrm{i}\omega})|^2/2\pi)\,\mathrm{d}\omega$. The Fourier expansion for the power spectrum $W = |H|^2$ is $\sum_k c_k t^{-k}$, $t \in \mathbb{T}$ where the c_k are the autocorrelation coefficients of the signal h. These coefficients satisfy $c_{-k} = \bar{c}_k$. Since $|H|^2 \geqslant 0$, the Toplitz matrix $T_W = [c_{i-j}]$ is positive definite. Note also that $\|A_*\|_\mu = \|A\|_\mu$ for any A and any positive measure on $\mathbb{T}$.

Once the predictor $A = F_*$ is known, we can invert the whitening filter relation $E = AH$ and model H as E/A. However, E is not known, but if the prediction is good, then we can assume that it has a relatively flat spectrum. If the infimum in the Szegő problem is $G^2 \geqslant 0$, then we can approximate H by $\hat{H} = G/A$. Thus, the signal h is approximated as the impulse response of a system with transfer function $\hat{H}$ which is called the modeling filter.

Szegő's theorem says that if $W \in L_1(\mathbb{T})$, then the optimal predictor is $\hat{H}$ with $\hat{H}_*(z) = F(z) := S(0)/S(z)$ where for $c \in \mathbb{T}$ and $z \in \mathbb{D}$

$$S(z) = c \exp\left\{ \frac{1}{4\pi} \int_{\mathbb{T}} D(t,z) \log W(t)\, \mathrm{d}\omega \right\}, \quad D(t,z) = \frac{t+z}{t-z}$$

is an outer spectral factor of W. If W is rational, then this means that S has no zeros and no poles in $\mathbb{D}$ and $|S(t)|^2 = W(t)$ for $t \in \mathbb{T}$.

The practical computation of this infinite-dimensional problem is by computing the solution of the prediction (alias Szegő) problem in finite-dimensional subspaces of H_2^μ. Let $\mathscr{L}_n$ be $(n+1)$-dimensional subspaces that are nested $\cdots \subset \mathscr{L}_n \subset \mathscr{L}_{n+1} \subset \cdots$ and such that the L_2^μ-closure of $\bigcup \mathscr{L}_n$ is H_2^μ. We then try to solve the partial Szegő problem

Problem 4.2. (Partial Szegő) *Find a function $F_n \in \mathscr{L}_n \subset H_2^\mu$, with $F_n(0) = 1$ such that we solve one of the following problems*:

- *minimum norm problem*: *given n, find F_n that minimizes $\|F_n\|_\mu$, or*
- *minimum degree problem*: *given ε, find F_n with smallest n such that $\|F_n\|_\mu \leqslant \varepsilon$.*

If $\{\phi_k\}_{k=0}^n$ is an orthonormal basis for $\mathscr{L}_n$, $n = 0, 1, \ldots$, then $k_n(z,w) = \sum_{k=0}^n \phi_k(z)\overline{\phi_k(w)}$ is a reproducing kernel for $\mathscr{L}_n$. This means that $\langle f, k_n(\cdot, w)\rangle_\mu = f(w)$ for all $f \in \mathscr{L}_n$. The solution F_n of the minimum norm partial Szegő problem in $\mathscr{L}_n$ is then given by $k_n(z,0)/k_n(0,0)$ and the infimum is $1/k_n(0,0)$.

In the original theory of Grenander, Wiener and Szegő, the subspaces were the polynomial subspaces: $\mathscr{L}_n = \Pi_n$, the polynomials of degree at most n. Then, it can be shown that $k_n(z,0) = \kappa_n \phi_n^\#(z) = \kappa_n z^n \phi_{n*}(z)$ with $\{\phi_n\}$ the orthonormal polynomials and $\kappa_n > 0$ the leading coefficient of ϕ_n. Thus $\kappa_n = \phi_n^\#(0)$ and so $k_n(0,0) = \kappa_n^2$, giving $\hat{H}_n(z) = z^n/\phi_n(z)$. Note that the zeros of ϕ_n are in $\mathbb{D}$ so that this is a stable and minimal phase transfer function, or in mathematical terms a conjugate outer function since all its poles and zeros are in $\mathbb{D}$. This solution also results in a *Chebyshev–Padé* approximation of $W(z)$ since indeed the Fourier series of $W_n = |\hat{H}_n|^2$ is $\sum_k \hat{c}_k \mathrm{e}^{-\mathrm{i}k\omega}$ where $\hat{c}_k = c_k$ for all $|k| < n+1$. Moreover, it can be shown that $\mathrm{d}\mu_n = (W_n/2\pi)\,\mathrm{d}\omega$, solves the partial trigonometric moment problem as formulated below.

Problem 4.3. (Trigonometric moments) *Given c_k, $k \in \mathbb{Z}$, find a positive measure on $\mathbb{T}$ such that $c_k = \int_{\mathbb{T}} t^k\, \mathrm{d}\mu(t)$, $k \in \mathbb{Z}$. The partial problem is*: *given $c_0, \ldots, c_n$, find a positive measure μ_n such that it has exactly these moments or equivalently such that $\langle f, g\rangle_\mu = \langle f, g\rangle_{\mu_n}$ for all $f, g \in \Pi_n$.*

Thus $\int_{\mathbb{T}} t^k\, \mathrm{d}\mu_n(t) = \int_{\mathbb{T}} t^k\, \mathrm{d}\mu(t)$ for $|k| < n+1$. The Riesz–Herglotz transform of μ, given by

$$C(z) = \int_{\mathbb{T}} D(t,z)\, \mathrm{d}\mu(t), \quad D(t,z) = \frac{t+z}{t-z}$$

is in the Carathéodory class $\mathscr{C} = \{C \in H_\infty\colon\ \mathrm{Re}\, C(z) \geqslant 0,\ z \in \mathbb{D}\}$, and it has the expansion $C(z) = c_0/2 + \sum_{k=1}^\infty c_k z^k$, $z \in \mathbb{D}$. The c_k are the trigonometric moments. If C_n is the Riesz–Herglotz transform of μ_n, then $C(z) - C_n(z) = \mathrm{O}(z^{n+1})$ for $z \to 0$, so that we have also solved a Carathéodory–Fejér interpolation problem (see below).

When the construction of the orthogonal polynomials are formulated in terms of linear algebra, then the coefficients of the orthogonal polynomials are obtained as the solution of the Yule–Walker equations, which are in fact the normal equations for the least-squares problem posed by the linear prediction. The matrix of the system is a (positive-definite Hermitian) Toeplitz matrix, and again, like in the Hankel case, fast algorithms exist that solve subsystems by considering the leading principal submatrices of the Toeplitz matrix. For the duality between Hankel and Toeplitz systems and the associated (formal) orthogonal polynomials, see [13]. However, here the Toeplitz matrices are positive definite, unlike the Hankel systems in the partial realization problem. Therefore, in the linear prediction problem we are not confronted with a numerical and system theoretical instability problem as in the partial realization problem.

The well-known Levinson algorithm is a fast (i.e. $O(n^2)$) algorithm to solve Toeplitz systems. It is a version of the Gram–Schmidt orthogonalization that is fast because it exploits the structure of the Toeplitz matrix. Since the solution of the system is computed recursively, the algorithm computes as a bonus the so-called *reflection coefficients* that are related to the recurrence coefficients that occurred in the recurrence relations for the orthogonal polynomials as derived by Szegő. They are also called Schur coefficients because they also occur in a continued fraction like algorithm that was designed by Schur to see whether a given function is in the Schur class, that is the class $\mathscr{S} = \{f \in H_\infty : \|f\|_\infty \leqslant 1\}$. The Schur algorithm is based on the simple lemma that $f_k \in \mathscr{S}$ iff $\rho_k = f_k(0) \in \mathbb{D}$ and $f_{k+1} = (1/z)[f_k - \rho_k]/[1 - \bar{\rho}_k f_k] \in \mathscr{S}$. These ρ_k are the Schur parameters. When translated in terms of linear algebra we can say that the Levinson algorithm gives a L^*DL factorization of T_W^{-1} while the Schur algorithm gives an LDL^* factorization of T_W where T_W is the Toeplitz matrix with symbol W. The Jury test to check the discrete stability of a polynomial (i.e., to see whether all its zeros are inside the unit circle) can also be seen as a variant of the Schur algorithm.

The mathematical ideas that we have just described were developed around the beginning of the 20th century. The multipoint generalization by Nevanlinna and Pick was published around 1920. The breakthrough of this multipoint generalization in signal processing, system theory, inverse scattering, transmission lines, etc., came not before the 1980s and was related to the introduction of the AAK ideas.

Let us reconsider the linear prediction problem, but now we take for $\mathscr{L}_n$ subspaces of rational functions to be defined as follows. Let $\{z_k\}_{k=1}^{\infty}$ be a sequence of not necessarily different points in $\mathbb{D}$ and set $\pi_0 = 1$ and $\pi_n(z) = \prod_{k=1}^{n}(1 - \bar{z}_k z)$. The spaces $\mathscr{L}_n$ are then defined as the rational functions of degree at most n whose denominator is π_n: $\mathscr{L}_n = \{f = p_n/\pi_n : p_n \in \Pi_n\}$. Defining the Blaschke factors $\zeta_k(z) = \eta_k(z - z_k)/(1 - \bar{z}_k z)$, $k = 1, 2, \ldots$ where $\eta_k = 1$ if $z_k = 0$ and $\eta_k = -\bar{z}_k/|z_k|$ otherwise, then it is obvious that $\mathscr{L}_n$ is spanned by the Blaschke products $B_0 = 1$ and $B_k = \prod_{i=1}^{k} \zeta_i$, $k = 1, \ldots, n$. Note that if we choose all $z_k = 0$, then we are back in the polynomial case.

Following the same lines as above, we can construct an orthogonal basis by Gram–Schmidt orthogonalization of the $\{B_k\}$. Let us denote the orthonormal basis as $\phi_k = a_{k0}B_0 + a_{k1}B_1 + \cdots + a_{kk}B_k$, with $\kappa_k = a_{kk} > 0$. To solve the partial linear prediction problem in $\mathscr{L}_n$, we then construct $k_n(z, 0)/k_n(0, 0)$ where $k_n(z, w) = \sum_{k=0}^{n} \phi_k(z)\overline{\phi_k(w)}$. However, in the general case $k_n(z, 0)$ will not simplify as in the polynomial case so that we are stuck with the expression $\hat{H} = 1/K_{n*}$ with $K_n(z) = k_n(z, 0)/\sqrt{k_n(0, 0)}$ but this is again a minimal phase and stable transfer function. Indeed, $\hat{H}_n$ is of the form $\prod_{k=1}^{n}(z - z_k)/P_n(z)$ where P_n is a polynomial with all its zeros in $\mathbb{D}$.

All the approximation properties that we had before are transformed into multipoint versions. For example, if $W_n = |\hat{H}_n|^2$ and $\mathrm{d}\mu_n = (W_n/2\pi)\,\mathrm{d}\omega$, then $\langle f, g\rangle_\mu = \langle f, g\rangle_{\mu_n}$ for all $f, g \in \mathscr{L}_n$. This means that we solve a partial moment problem in $\mathscr{L}_n \cdot \mathscr{L}_{n*}$ where the moments are given by $c_k = \int_{\mathbb{T}} B_k(t)\,\mathrm{d}\mu(t)$, $k \in \mathbb{Z}$ with $B_{-k} = B_{k*} = 1/B_k$. The Schur interpolation algorithm is replaced by the Nevanlinna–Pick (NP) interpolation algorithm, which solves the NP interpolation problem.

Problem 4.4. (Nevanlinna–Pick interpolation) *Given* $z_0 = 0, z_1, z_2, \ldots, z_n \in \mathbb{D}$, *and* $w_0 = 0, w_1, w_2, \ldots, w_n \in \mathbb{C}$, *find a function* $F \in H_\infty$ *such that* $F(z_k) = w_k$, $k = 0, 1, 2, \ldots, n$. *For the partial problem*, n *is finite*, *for the full problem*, n *is infinite. If there are more solutions*, *one can characterize all the solutions and solve one of the following two problems*:

- *minimum norm problem*: *find a solution with minimal norm* $\|F\|_\infty$ *or the*
- *minimum degree problem*: *among the solutions with* $\|F\|_\infty < \varepsilon$, *find one of minimal degree.*

This formulation corresponds to mutually different points z_i, but it is not difficult to imagine what the confluent case involves. If several of the z_k points coincide, then it is, in fact, a reformulation of an Hermite–Fejér problem and if all the points coincide at $z_k = 0$, then the NP algorithm becomes the Schur algorithm and the above problem becomes a Schur or Carathéodory–Fejér problem, although the latter is usually formulated by conformally mapping the function values from the unit circle to the right half-plane, so that the Schur class $\mathscr{S}$ is replaced by the Carathéodory class $\mathscr{C}$. The original NP and Schur algorithms just checked whether some F was a Schur function, thus whether there is a solution with $\|F\|_\infty \leqslant 1$. Like the Schur algorithm, the NP algorithm is based on a simple lemma that is a slight generalization of the Schur lemma: $f_k \in \mathscr{S}$ iff for some $z_k \in \mathbb{D}$, $\rho_k = f_k(z_k) \in \mathbb{D}$ and $f_{k+1} = (1/\zeta_k)[f_k - \rho_k]/[1 - \bar{\rho}_k f_k] \in \mathscr{S}$.

This is a good place to introduce the *Nehari problem* since it can be seen as a generalization of the NP problem and hence also of the Schur problem.

Problem 4.5. (Nehari) *Given* $\{h_k\}_{k=1}^\infty$, *find the function* $H \in L_\infty$ *such that* $\|H\|_\infty$ *is minimal and* $h_k = (1/2\pi)\int_0^{2\pi} H(\mathrm{e}^{\mathrm{i}\omega})\mathrm{e}^{\mathrm{i}k\omega}\,\mathrm{d}\omega$, $k = 1, 2, \ldots$.

If we define B_n as the Blaschke product with zeros $z_1, \ldots, z_n$ and $G \in H_\infty$ as a function that satisfies the partial NP interpolation conditions, then the set of all functions in H_∞ satisfying the partial NP interpolation conditions is given by $G + B_n H_\infty$, and a minimal norm solution F is given by solving the Nehari problem $\inf\{\|B_n^{-1}G - H\|_\infty : H \in H_\infty\}$ and setting $F = G - B_n H$. This minimal norm solution is rational. As in previous problems, one can require that (after rescaling) $\|F\|_\infty \leqslant 1$, and then find the rational solution with minimal degree.

The Nehari problem is particularly important for applications in control theory, where it usually appears under a slightly modified form which is a minimum degree version: find a solution satisfying $\|H\|_\infty \leqslant \varepsilon$ for a given ε. By appropriate rescaling, one can reduce this problem to the standard form where $\varepsilon = 1$. If there is more than one solution then one could select the one with the minimal degree.

The relation between the Nehari problem and the minimal norm AAK problem should be clear. Given a Hankel operator Γ_H, i.e., the numbers $\{h_k\}_{k=1}^\infty$, find a function $E \in L_\infty$ such that $h_k =$

$(1/2\pi)\int_0^{2\pi} E(e^{i\omega})e^{ik\omega}\,d\omega$ and such that it has minimal norm $\|E\|_\infty = \sigma_{n+1}$. The latter means that E is equal to σ_{n+1} times an all pass function, or in mathematical terms, a function that is unimodular on $\mathbb{T}$ almost everywhere, and the solution of the AAK problem is then $\hat{H} = H - E$. Note that if H is rational, then E is rational as well. Thus the all pass function is a Blaschke product and thus is E equal to σ_{n+1} times a Blaschke product.

5. Chain scattering matrices and H_∞ control

Consider a 2×2 matrix valued functions $M(z)$, and suppose J is a constant diagonal matrix with diagonal entries $+1$ and -1. Then M is called J-unitary if $M(z)^* J M(z) = J$. The matrix is called J-contractive if $J - M(z)^* J M(z) \geqslant 0$ where the inequality is to be understood in the sense of positive-definite matrices.

In a more general context, J-unitary matrices were studied by Potapov [43]. It was only since 1978 [24,21,22] that Dewilde and coworkers used the full power of J-unitary matrices in prediction theory and related matters like transmission lines, and inverse scattering. In fact, the recurrence relation for the orthogonal rational functions ϕ_n and for the reproducing kernels $k_n(z,w)$, of the previous section can be written in the form of a matrix relation. For example with $K_n(z) = k_n(z,0)/\sqrt{k_n(0,0)}$ and $K_n^\#(z) = B_n(z)K_{n*}(z)$, there exist J-unitary matrices $\theta_n(z)$ such that

$$\begin{bmatrix} K_{n+1}^\#(z) \\ K_{n+1}(z) \end{bmatrix} = \theta_n(z) \begin{bmatrix} K_n^\#(z) \\ K_n(z) \end{bmatrix} = \theta_n(z)\cdots\theta_0(z) \begin{bmatrix} K_0^\#(z) \\ K_0(z) \end{bmatrix} = \Theta_n(z) \begin{bmatrix} 1 \\ 1 \end{bmatrix},$$

where we assumed an appropriate normalization: $\int_{\mathbb{T}} d\mu(t) = 1$, so that $\phi_0 = 1$ and hence $K_0 = K_0^\# = 1$. Since the product of J-unitary matrices is a J-unitary matrix, the matrix Θ_n represents a scattering medium consisting of n layers. At the boundary of layers n and $n+1$, K_n and $K_n^\#$ can be considered as incident and reflected wave on the n side of the boundary, while at side $n+1$ of the boundary, we have K_{n+1} and $K_{n+1}^\#$. The θ_n are called *chain scattering matrices* (CSM) because if the medium consists of several layers, then the CSM for the whole medium is the product of the CSMs of each layer. Adding one extra layer just requires an extra θ factor. Since the matrix θ_n will depend on the part of the energy that is reflected and the part that is transmitted, it will depend on the reflection coefficients. In fact, this is the origin of the name reflection coefficient. The variable z enters as a delay operator representing the time needed for the wave to pass through and back an homogeneous layer. Physically, if the system is passive, i.e., if it does not add or absorb energy, then the CSM is J-unitary in $\mathbb{T}$ and J-contractive in $\mathbb{D}$. It also explains why the reflection coefficients are bounded by 1 in modulus: they represent the fraction that is reflected.

In terms of electrical circuits [5], the θ matrices represent a 2-port (two I/O pairs) mapping one I/O pair into another I/O pair. A CSM is equivalent to a scattering matrix mapping inputs into outputs. A scattering matrix of a passive network is a unitary matrix on $\mathbb{T}$ and contractive in $\mathbb{D}$, but the concatenation of 2-ports gives rise to a complicated star product for the scattering matrices, replacing the ordinary product of the CSMs.

The special structure of the θ matrices does not only give a direct lattice realization of the whitening filter (analysis) or modeling filter (synthesis), but they can even be used for the design of dedicated hardware implementation with systolic arrays [36].

The CSM can also play a prominent role in H_∞ control. Let us consider a simple standard control problem

$$\begin{bmatrix} Z \\ Y \end{bmatrix} = \begin{bmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{bmatrix} \begin{bmatrix} W \\ U \end{bmatrix},$$
$$U = KY,$$

where Z is the errors to be controlled, W the exogenous input, Y the observed output, U the control input, P the plant matrix, K the controller.

The closed-loop transfer function from W to Z is

$$H = P_{11} + P_{12}K(I - P_{22}K)^{-1}P_{21}.$$

Thus if W is some disturbing noise, then the controller should ensure that the observed error Z is only influenced marginally.

Problem 5.1. (H^∞ control) *Find the controller K such that $\|H\|_\infty < \gamma$ and such that the system is internally stable.*

Internal stability means that no fuses in the plant may burn, i.e., no state variable in the plant should become infinite. This problem can be reformulated in terms of CSMs. Indeed, the natural way to look at this problem is to consider the plant P as a 2-port (inputs W and U, outputs Z and Y) that is loaded by the controller: a 1-port (input Y, output U). The above description is a typical scattering matrix formulation, mapping inputs into outputs. However a CSM approach is much more interesting. Then the I/O pair (Y, U) is mapped into the I/O pair (W, Z). This gives

$$\begin{bmatrix} Z \\ W \end{bmatrix} = \Theta \begin{bmatrix} U \\ Y \end{bmatrix} \quad \text{with} \quad \begin{bmatrix} Z \\ W \end{bmatrix} = \begin{bmatrix} H \\ I \end{bmatrix} W \quad \text{and} \quad \begin{bmatrix} U \\ Y \end{bmatrix} = \begin{bmatrix} K \\ I \end{bmatrix} Y.$$

Note that this relation expresses that H is a linear fractional transform of K, characterized by Θ: $H = \mathrm{LFT}(\Theta; K)$. If Θ is J-unitary, then it maps $K \in \mathscr{S}$ into $H \in \mathscr{S}$. Thus $\|H\|_\infty < \gamma$ if $\|K\|_\infty < \gamma$. The solvability of the control problem reduces to the existence of a J-lossless factorization. This is an indefinite matrix version of the classical inner-outer factorization of complex functions. Assume that Θ is the CSM representation of a plant and assume it can be factorized as $\Theta = \Delta\Pi$ with Π an invertible J-unitary matrix and Δ is the CSM of a stable plant, which means that it is J-contractive in $\mathbb{D}$. Thus Π absorbs the instabilities of the plant. Since Δ now has the property that any controller $\tilde{K}$, with $\|\tilde{K}\|_\infty < \gamma$ will solve the problem, we have an infinite set of controllers for the plant Θ given by $K = \mathrm{LFT}(\Pi^{-1}; \tilde{K})$ since indeed $H = \mathrm{LFT}(\Theta; K) = \mathrm{LFT}(\Theta\Pi^{-1}; \mathrm{LFT}(\Pi; K)) = \mathrm{LFT}(\Delta; \tilde{K})$. This J-lossless factorization can be obtained by a Nevanlinna–Pick type of algorithm. For more details we refer to [38] where it is also shown that many other related control problems can be reduced to the present one or a vector generalization thereof.

6. Identification

We consider an identification problem in the frequency domain. Suppose we know the frequency response H at some specific points $\{t_i : i = 1, \ldots, N\} \subset \mathbb{T}$. Depending on what norm or what other objectives one wants to achieve, there are several different approximation problems to solve.

6.1. Discrete linearized least squares

As a first example, one can try to find a rational function (of smallest possible degree) that interpolates the given data, or if a restriction is given on the degree, find the one that interpolates in a discrete least-squares sense. The problem remains linear if we multiply out the denominator.

Problem 6.1. (Discrete linearized least squares) *Given* $\{t_i\}_{i=1}^N \subset \mathbb{T}$ *and couples of numbers* $\{(f_i, e_i)\}_{i=1}^N$ *such that* $H(t_i) = f_i/e_i$, $i = 1, \ldots, N$, *find an approximant* $\hat{H}(z) = N(z)/D(z)$ *of degree at most n such that* $\sum_{i=1}^N w_i |R_i|^2$ *is minimal where* $R_i = N(t_i)e_i - D(t_i)f_i$ *and* $\{w_i\}_{i=1}^N$ *are positive weights.*

The solution of this problem is again related to the Schur algorithm, but now with a discrete measure that has some mass concentrated only at the points t_i. For such a measure the Schur algorithm will break down after a finite number of steps because a reflection coefficient will become one in modulus. If the degree n is reached before all the interpolation conditions are satisfied, then the algorithm is artificially forced to an end, pushing some masses to the points t_i on $\mathbb{T}$. Linear least-squares problems are typically solved using orthogonal polynomials and can therefore be formulated solely in terms of linear algebra. A fast algorithm for discrete polynomial least-squares approximation on the real line is given by Forsythe (1957). It performs a Gram–Schmidt orthogonalization and stores the polynomials by their three-term recurrence relation. The analog for the unit circle uses the Szegő recurrence and stores the polynomials via their reflection coefficients [46]. More precisely, we write the solution as a polynomial vector

$$S(z) = \begin{bmatrix} D(z) \\ N(z) \end{bmatrix} = \sum_{i=0}^{n} \Phi_i(z) C_i, \quad C_i = \begin{bmatrix} \rho_i \\ \tau_i \end{bmatrix},$$

where the C_i are constant vectors and Φ_i are 2×2 polynomial matrices that are orthogonal in the following sense:

$$\langle \Phi_k, \Phi_l \rangle_W = \sum_{i=1}^{N} \Phi_k^*(t_i) W_i \Phi_l(t_i) = \delta_{k,l} I_2, \quad W_i = E_i^* v_i E_i, \quad E_i = [f_i \quad -e_i].$$

The superscript $*$ denotes the adjoint. We want to minimize $\langle S(z), S(z) \rangle_W$. The minimum is obtained for all $C_i = 0$, except for C_n, because the degree should be n. Choosing the two variables ρ_n and τ_n is a very simple problem because $\langle S(z), S(z) \rangle_W = |\rho_n|^2 + |\tau_n|^2$. The algorithm to generate the orthogonal polynomials Φ_i is a block version of a Szegő-like recurrence relation. It results in the following recursive matrix interpretation. Define the matrix E whose ith row is E_i, $i = 1, \ldots, N$ and $Z = \operatorname{diag}(t_1, \ldots, t_N)$. Then define

$$M = \left[\begin{array}{c|c} 0 & E^* \\ \hline E & Z \end{array}\right] \quad \text{and} \quad \tilde{M} = Q^* M Q = \left[\begin{array}{c|cc} 0 & \sigma_0^* & 0 \\ \hline \sigma_0 & & \\ 0 & & H \end{array}\right],$$

where Q is a unitary matrix, σ_0 is a 1×2 vector and H is a unitary upper Hessenberg matrix. It is unitary because it is a unitary similarity transformation of the original matrix Z, which is unitary. If $\tilde{M}$ is extended with the new data E_{N+1} and t_{N+1}, then the same Hessenberg structure is restored

by a unitary similarity transformation. This updating is fast because the Hessenberg matrix is stored by its block Schur parameters, and this update is performed very efficiently.

6.2. Robust identification

Robustness is an important issue in systems and control. It essentially means that it is not only sufficient to solve the problem within some tolerance, but the performance should remain within this tolerance in the worst-case situation when certain perturbations are allowed. For example, in the control problem we had to design a controller that generated a transfer H with $\|H\|_\infty < \gamma$. In *robust control*, this should be satisfied for all possible perturbations of the plant P that remain within certain bounds. Sometimes this robustness is obtained by formulating a weighted problem where the weight is chosen so as to emphasize the sensitive parts of the objective function. There is also a technique of structured singular values or μ-analysis where some singular value is monitored within some bounded perturbations of variables that cause a structured perturbation of the matrix. Because this is essentially a linear algebra technique, we will not go deeper into this matter. As an example of a robustness problem, we discuss here the robust identification problem.

Suppose that we know the frequency response H of a stable system in N points $t_i \in \mathbb{T},\ i = 1,\dots,N$. We can, of course, find an approximant $\hat{H}_N$ and we want the algorithm to be such that $\hat{H}_N \to H$ for $N \to \infty$ for all stable H. By a mapping $z \mapsto z^{-1}$, we can reformulate the problem in the disk algebra $\mathscr{A}$. Thus setting $F = H_*$, we have $F \in \mathscr{A}$. Robustness means that we now allow the observed data to be contaminated by noise. Thus, we are given $F(t_i) + \eta_i$ with $|\eta_i| < \varepsilon$ (i.e., $\|\eta\|_\infty < \varepsilon$). We still want the algorithm to be such that in the worst case situation, the convergence still holds for $\varepsilon \to 0$. The problem is thus described as follows.

Problem 6.2. (Robust identification) *Given is a function $F \in \mathscr{A}$. This function can be computed in points $t_i \in \mathbb{T}$ with some error η_i bounded by ε, so that we can compute $\tilde{F}_i := F(t_i) + \eta_i,\ i = 1,\dots,N$ with $t_i \in \mathbb{T}$ and $\|\eta\|_\infty < \varepsilon$. Design an algorithm $A_N : \mathscr{A} \to \mathbb{C}^N \to \mathscr{A} : F \mapsto \hat{F}_N$ that constructs an approximant $\hat{F}_N$ using the values $\{\tilde{F}_i\}_{i=1}^N$ such that*

$$\lim_{N \to \infty, \varepsilon \to 0} \sup_{\|\eta\|_\infty < \varepsilon} \sup_{F \in \mathscr{A}} \|\hat{F}_N - F\|_\infty = 0.$$

It was shown by Partington in 1992 that there is no linear algorithm that solves this problem. So the problem is usually solved in two steps.

(1) Find a *linear* algorithm V_N that constructs an approximant G_N in L_∞.
(2) Find a rational approximant $\hat{F}_N \in \mathscr{A}$ of $G_N \in L_\infty$ that minimizes $\|G_N - \hat{F}_N\|_\infty$.

The first problem is a problem of sampling theory: how much (i.e., what samples) do we need to know about a function to be able to recover it, if we know that the function is in a certain class. The second problem is in fact a Nehari problem, that has been discussed before (Problem 4.5). Suppose the approximant of F generated by V_N, given $\tilde{F}_i = F(t_i) + \eta_i,\ i = 1,\dots,N$, is denoted as $V_N(F + \eta)$. Because V_N is linear and $\|\eta\|_\infty < \varepsilon$, we have

$$\begin{aligned} \|V_N(F+\eta) - F\|_\infty &\leqslant \|V_N(F) - F\|_\infty + \|V_N(\eta)\|_\infty \\ &\leqslant \|V_N(F) - F\|_\infty + \|V_N\|_\infty \varepsilon. \end{aligned}$$

Thus $V_N(F+\eta)$ will converge to F in the sense of robust identification in the set L_∞ if $\lim_{N\to\infty} \|V_N(F)-F\|_\infty = 0$ and V_N is bounded in $\mathscr{A}$.

A simple example for the algorithm V_N is to choose t_l as the Nth roots of unity $t_l = e^{2\pi i l/N}$ and to compute the discrete Fourier transform (DFT) for the given samples: $c_N(k) = (1/N)\sum_{l=1}^{N} f(t_l)t_l^{-k}$, $k\in\mathbb{Z}$, which defines an approximant $\tilde{G}_{n,N}(e^{i\omega}) = \sum_{k=-n}^{n} c_N(k)e^{ik\omega}$. However, this approximant does not in general converge to F when n and N go to infinity. We need a special summation technique, for example the Fejér sum, which takes the average of the $\tilde{G}_{k,N}$, $k=1,\dots,n$: $G_{n,N} = (1/n)\sum_{k=1}^{n} \tilde{G}_{k,N}$. This corresponds to the introduction of weight coefficients in the DFT, thus we use a windowed DFT: $G_{n,N}(e^{i\omega}) = \sum_{k=-n}^{n} w_k c_N(k)e^{ik\omega}$ with $w_k = 1-|k|/n$. Other summation techniques (e.g., de la Vallée-Poussin) exist that correspond to other weights. Finally, the approximant G_N is then given as $G_N = \lim_{n\to\infty} G_{n,N}$.

This construction of trigonometric approximants should be generalized so that we need not necessarily take equidistant points on $\mathbb{T}$. However, it is intuitively clear (and a hard proof exists) that the points t_i should be eventually dense in $\mathbb{T}$ if we want to recover F exactly. Of course, modifications can be made if we know that the signal is band limited, or if only approximants in a certain band are important, while outside that interval, the approximant may even diverge.

Anyway, when using trigonometric approximants, the convergence may be very slow, especially when there are poles close to $\mathbb{T}$. In that case it might be wise to use rational basis functions instead of (trigonometric) polynomials. In fact, it was for a similar reason that in prediction theory, the Nevanlinna–Pick algorithm replaced the Schur algorithm, so that AR models could be replaced by ARMA models.

Assume that we have some estimates of the poles. So we are given a sequence of (not necessarily distinct) $z_k\in\mathbb{D}$. The orthogonal rational functions ϕ_n with poles $\{1/\bar{z}_k\}_{k=1}^{n}$ as discussed in Section 4, but now made orthogonal with respect to the Lebesgue measure on $\mathbb{T}$ are known explicitly (Walsh attributes them to Malmquist) and they are given by

$$\phi_n(z) = \frac{\sqrt{1-|z_n|^2}}{1-\overline{z_n}z}\, z\prod_{k=1}^{n-1}\frac{z-z_k}{1-\bar{z}_k z}, \quad n\geqslant 0.$$

It is known that $\operatorname{span}\{\phi_k: k=0,1,\dots\}$ is dense in H_2 and in the disk algebra $\mathscr{A}$ iff $\sum_{k=0}^{\infty}(1-|z_k|)=\infty$. Some special cases became rather popular in identification: when all the z_k are equal to some fixed $a\in[-1,1]$, then this system is known as the Laguerre system because it is related to a transformation of the Laguerre functions. This system was used by Wahlberg in 1991. In 1994 he also introduced the Kautz system based an a complex conjugate pair: $z_k = \bar{z}_{k+1} = a$. For a survey of all kinds of orthogonal (rational) bases and their use in system identification see [41]. Schipp and Bokor proposed yet another system which corresponds to a cyclic repetition of the poles $\{1/\bar{z}_1,\dots,1/\bar{z}_d\}$. It is constructed as follows. Let B_d be the Blaschke product with zeros $z_1,\dots,z_d$ and consider an orthonormal basis $\{\phi_1,\dots,\phi_d\}$ of $B_dH_1(\mathbb{D})$. Then $\{\phi_l B_d^k: 1\leqslant l\leqslant d; k=0,1,\dots\}$ is an orthonormal basis for $H_2(\mathbb{D})$.

More recently, also rational wavelet-like bases were used to represent functions in the disk algebra. For example, the Franklin system is a piecewise linear L_2-orthogonal system in $C(\mathbb{T})$. First let $\psi(\omega)$ be the hat function in $[0,\pi]$ (zero in 0 and π, one in $\pi/2$ and linear in between). Then define $\psi_{nk}(e^{i\omega}) = \psi(2^n\omega - k\pi)$ for $k=0,\dots,2^n$ and $n\in\mathbb{N}$. These functions are orthogonalized to give a Faber–Schauder basis ϕ_{nk} for functions in $C(\mathbb{T}^+)$ where $\mathbb{T}^+$ means the upper half of $\mathbb{T}$. Because the trigonometric conjugate system $\tilde{\phi}_{nk}$ is also continuous, the functions $\Phi_{nk} := \phi_{nk} + i\tilde{\phi}_{nk}$, when

extended to the lower half of $\mathbb{T}$ by $\Phi_{nk*} = \Phi_{nk}$ and analytically extended to $\mathbb{D}$, form a basis for the disk algebra $\mathscr{A}$.

Another rational wavelet-like basis can be obtained by using the the Cauchy kernel $C(z,w) = 1/(1-\bar{w}z)$. Therefore define the set $W = \{w_{nk} = \rho_n e^{i\omega_{nk}}: k = 0,\ldots,2^n - 1; n = 1,2,\ldots\}$ where $\rho_n = 1 - 2^{-n}$ and $\omega_{nk} = 2k\pi 2^{-n}$. The system $\{C(z,w): w \in W\}$ is dense in the disk algebra $\mathscr{A}$. In fact, this is an overcomplete system because it is sufficient that $\sum_{k,n}(1 - |w_{nk}|) = \infty$ for the system to be dense.

7. Generalizations: past and future

The basic idea relating *Padé-like approximations* and *partial realization* has been extended in many directions, e.g., to MIMO systems (giving rise to block Hankel matrices) and to two-point Padé approximations, (using information at 0 and ∞) and multipoint Padé approximation and general rational interpolation. Sometimes a combination of Markov parameters and correlation coefficients are fitted like in q-COVER [44]. Many generalizations of the Padé approximation and interpolation problem lead to new developments Laurent–Padé [7], vector-Padé [2]. Many *matrix valued rational interpolation* problems are for example discussed in [4] and related problems are discussed in several volumes of the same series, edited by Gohberg. It is also interesting to see how many techniques like state-space descriptions, controllability and observability matrices, are used in the analysis and solution of these problems. See [11] for a bibliography on Padé techniques in systems theory. The original matrix minimal partial realization problem got a first satisfactory solution in state space from, described by Ho and Kalman [35].

In many applications, notably in controller design applications, the objective of the model reduction problem can be weighted in the frequency domain. The reason is that often (especially in control system design) one is interested in a good match between the reduced model and the original one, at a certain frequency or in the neigborhood of a certain frequency (e.g. the so-called *cross-over* frequency in control system design). Therefore typically frequency-domain weighting matrices are included in the model reduction framework, so as to minimize the input–output weighted error $W_o(z)[H(z) - \hat{H}(z)]W_i(z)$. Here $W_o(z)$ and $W_i(z)$ are the output, resp. input weighting functions that emphasize certain frequency regions in which the approximation error should preferably be small. References to extensions in this direction to H_2-optimal model reduction, Hankel norm approximation and balanced truncation can be found in [44, p. 33]. An important special case occurs when one of the weights in the weighted error is the inverse of the original system, in which case one is minimizing the so-called *relative error*. Balanced stochastic truncation is one such method that achieves balanced truncation in a relative error framework (see [44, Chapter 4] for a survey and some error bounds upper bounding the H_∞-norm of the 'relative' error to the so-called balanced stochastic singular values). For references on frequency weighted open- and closed-loop balanced truncation, we refer to [44, Chapter 3].

The generalization of the realization problem to the situation where only input–output data of the system are available, $u_k, y_k,\ k = 0,1,2,\ldots$ (e.g., via measurements obtained from sensor devices), can be solved via so-called *prediction error methods* [39] or *subspace system identification methods* [47] (also check these references for literature surveys). Explaining this in detail would lead us too far. Suffice it to say that the identification problem is very important in many mathematical engineering problems in the process industry and that it leads to model-based control system design and optimization, softsensors, observers, etc.

Stimulated by the *AAK problem* from system theory, several related research projects were started in approximation theory. The fact that a Hankel norm approximation is, under certain conditions, nearly a best approximant in ∞-norm, was exploited by Trefethen and Gutknecht to construct near best polynomial and rational approximants. They called it CF approximation because they started from the classical *Carathéodory–Fejér* theorem that deals with the polynomial case (i.e., $n=0$). The system theoretic approximant is again diagonal in the sense that the degree of numerator and denominator are equal. This has been generalized in many directions including an analysis of the structure of a CF table in analogy with the Padé table [33]. The equi-oscillation condition of the error curve in real Chebyshev approximation on an interval is replaced by the circularity of the error curve and the winding number is large enough. As it is usually assumed in computations, H is a rational of high degree and $\hat{H}$ is a low degree approximant. Then the error $H-\hat{H}$ is σ_{n+1} times a Blaschke product. The winding number is then associated with the degree of this Blaschke product and this is related to the multiplicity of σ_{n+1} as a singular value [45].

There is a tendency to generalize one or two point interpolation (as in Padé approximation or in the Levinson algorithm) to more general *multipoint* (Hermite) *interpolation* problems (like in the Nevanlinna–Pick interpolation). This implies a shift from polynomial bases to rational bases. We gave some examples of this idea before. Here is another one. Recall that in the Nevanlinna–Pick problem, some points z_k were chosen which were used as interpolation points, but at the same time they featured as transmission zeros in the model. Recently, the problem was raised whether it is possible to keep all the nice properties of this approach but separate the role of interpolation points and transmission zeros [15]. The problem reduces to a constrained optimization problem to guarantee stability and yet obtain an interpolant that is of minimal degree. The search space is the set of all stable solutions to the interpolation problem which can, for example, be parametrized by the reflection coefficients.

The H_∞ *control problem* has many facets and many different kind of subproblems. The discussion given in Section 5 is just a start. It should be obvious that the analysis of *J-unitary matrices* is essential. The fine structure for matrix and operator valued functions that are unitary and/or contractive with respect to an indefinite matrix was initiated by Potapov, but because it is so essential in all kind of generalizations of the Nevanlinna–Pick theory, of moment problems, and all the engineering applications, the study of these matrices and all the related problems has grown out into an independent discipline generally known as *Schur analysis*. Some generalizations of Potapov's work are found in [27]. The importance of J-unitary matrices is surveyed in [10] which contains many references, especially to the Russian literature. For all kind of generalizations of the Nevanlinna–Pick algorithm see also [16].

The linear algebra problems (Hankel matrices in realization and Toeplitz matrices in linear prediction) were solved by fast algorithms because the structure of these matrices could be exploited. When considering signals which are not stationary, then the covariance matrix is not Toeplitz, but if the nonstationarity is only mild, then the matrices do not deviate too much from a Toeplitz matrices. The structural deviation from a Toeplitz matrix could be measured by the *displacement rank*, introduced around 1979 and studied by Kailath and many others [34,37]. Designing stable and fast or superfast algorithms for all sorts of structured matrices is still an active field of research.

The *nonstationary prediction* problem has been generalized to the time-varying case. The easiest way to see what this means is to consider the state-space description with the matrices (A,B,C,D). The *time-varying* case is obtained when we let these matrices depend on the (time) variable k

[3,23,31]. Many of the concepts, from Hankel matrix to orthogonal polynomials have been generalized. There is however much more work to be done before these techniques will be widely accepted.

The algorithm of Section 6.1 is a unit circle analogue for discrete time systems of an algorithm that starts from data given on the real line and that is used for continuous time systems. In the latter case the problem is related to a Hankel matrix and because of the symmetry, the (block) Hessenberg matrix will be (block) tridiagonal. This idea was first elaborated by Ammar and Gragg in 1984. Further generalizations are available and many variants of the algorithm exist and several applications are found in different domains. Also the so-called UHQR (unitary Hessenberg QR) and the TQR (tridiagonal QR) algorthms are closely related. They solve not the least squares, but the eigenvalue problems for unitary Hessenberg or tridiagonal matrices. Many references related to this circle of ideas are found in [28].

There is much research still going on about the generalization of classical *sampling theorems*. This is also stimulated by the interest in wavelets. The ultimate question is: how much function values does one have to know of a function to be able to reconstruct the function perfectly. The use of other bases than the usual complex exponentials is thereby an essential element. Mathematically, this is related to density problems: will the span of a certain set of basis functions or a frame be dense in the space where we want to approximate? There are many recent results, but there are still a lot of questions to be answered.

As for the use of *orthogonal rational bases*, there is a lot of theory accumulated on orthogonal rational functions with respect to a general positive measure on the unit circle (and on $\mathbb{R}$) in [9]. This monograph gives convergence and density results, interpolation properties and generalized moment problems in the scalar case. It is very useful if the ideas of identification are used with weighted approximation conditions, thus using another basis than the one orthogonal with respect to the Lebesgue measure. Moreover, this theory can be generalized to the matrix case and although there are many partial results available in the literature, a systematic study is not published. Another way of potential application of the monograph lies in the fact that also orthogonal rational functions are discussed whose poles are on the unit circle (and not inside the open disk $\mathbb{D}$). This has many potential applications in systems and identification that has not been explored so far.

The use of *wavelets* in identification is just starting up and a lot of work has to be done here. For example, the orthogonal rational wavelets based on reproducing kernels [8] may open a new horizon. Among the problems that are only partially explored, one can count the problem of selecting a best basis in a dictionary of possible bases.

An extremely useful survey from an approximation theoretical point of view about rational approximation and system related problems is given in [42].

References

[1] V.M. Adamjan, D.Z. Arov, M.G. Kreĭn, Analytic properties of Schmidt pairs for a Hankel operator and the generalized Schur–Takagi problem, Math. USSR-Sb. 15 (1971) 31–73.

[2] G.A. Baker Jr., P.R. Graves-Morris, Padé Approximants, 2nd Edition, Encyclopedia of Mathematics and its Applications, Vol. 59, Cambridge University Press, Cambridge, 1996.

[3] J.A. Ball, I. Gohberg, M.A. Kaashoek, Nevanlinna–Pick interpolation for time-varying input–output maps, in: I. Gohberg (Ed.), Time-Variant Systems and Interpolation, Vol. 56, Birkhäuser, Basel, 1992, pp. 1–51.

[4] J.A. Ball, I. Gohberg, L. Rodman, Interpolation of rational matrix functions, Operator Theory: Advances in Applications, Vol. 45, Birkhäuser, Basel, 1990.
[5] V. Belevitch, Classical Network Theory, Holden-Day, San Francisco, 1968, pp. 93,136,141. (Zbl 172.204).
[6] C. Brezinski, Padé-type approximation and general orthogonal polynomials, Operator Theory: Advances in Applications, International Series of Numerical Mathematics, Vol. 50, Birkhäuser, Basel, 1980.
[7] A. Bultheel, Laurent series and their Padé approximations, Operator Theory: Advances in Applications, Vol. OT-27, Birkhäuser, Basel, 1987.
[8] A. Bultheel, P. González-Vera, Wavelets by orthogonal rational kernels, in: B. Berndt, F. Gesztesy (Eds.), Continued Fractions: From Number Theory to Constructive Approximation, Contemporary Mathematics of the American Mathematical Society 236, (1999) 101–123.
[9] A. Bultheel, P. González-Vera, E. Hendriksen, O. Njåstad, in: Orthogonal Rational Functions, Cambridge Monographs on Applied and Computational Mathematics, Vol. 5, Cambridge University Press, Cambridge, 1999.
[10] A. Bultheel, K. Müller, On several aspects of J-inner functions in Schur analysis, Bull. Belgian Math. Soc. Simon Stevin 5 (1998) 603–648.
[11] A. Bultheel, M. Van Barel, Padé techniques for model reduction in linear system theory, J. Comput. Appl. Math. 14 (1986) 401–438.
[12] A. Bultheel, M. Van Barel, Linear prediction: mathematics and engineering, Bull. Belgian Math. Soc. Simon Stevin 1 (1994) 1–58.
[13] A. Bultheel, M. Van Barel, Formal orthogonal polynomials and Hankel/Toeplitz duality, Numer. Algorithms 10 (1995) 289–335.
[14] A. Bultheel, M. Van Barel, Linear Algebra, Rational Approximation and Orthogonal Polynomials, Studies in Computational Mathematics, Vol. 6, Elsevier, Amsterdam, 1997, 446 pp.
[15] C.I. Byrnes, S.V. Gusev, A. Lindquist, A convex optimization approach to the rational covariance extension problem, SIAM J. Control Optim. 37 (1) (1999) 211–229.
[16] T. Contantinescu, Schur Analysis, Factorization and Dilation Problems, Operator Theory: Advances and Applications, Vol. 82, Birkhäuser, Basel, 1996.
[17] B. De Moor, Structured total least squares and L_2 approximation problems, Linear Algebra Appl. 188–189 (1993) 63–207.
[18] B. De Moor, Total least squares for affinely structured matrices and the noisy realization problem, IEEE Trans. Signal Process. 42 (11) (1994) 3104–3113.
[19] B. De Moor, The Riemannian singular value decomposition, in: SVD and Signal Processing, III: Algorithms, Architectures and Applications, Elsevier, Amsterdam, 1995, pp. 61–78.
[20] B. De Moor, Structured total least squares for Hankel matrices, in: A. Paulraj, V. Roychowdhury, C. Schaper (Eds.), Communications, Computation, Control and Signal Processing, Kluwer, Dordrecht, 1997, pp. 243–258 (At the occasion of Thomas Kailath's 60th birthday, Stanford, USA).
[21] P. Dewilde, H. Dym, Schur recursions, error formulas, and convergence of rational estimators for stationary stochastic sequences, IEEE Trans. Inform. Theory IT-27 (1981) 446–461.
[22] P. Dewilde, H. Dym, Lossless inverse scattering, digital filters, and estimation theory, IEEE Trans. Inform. Theory IT-30 (1984) 644–662.
[23] P. Dewilde, A.J. Van Der Veen, Time-Varying Systems and Computations, Kluwer, Dordrecht, 1998.
[24] P. Dewilde, A. Viera, T. Kailath, On a generalized Szegő–Levinson realization algorithm for optimal linear predictors based on a network synthesis approach, IEEE Trans. Circuits Systems CAS-25 (1978) 663–675.
[25] R.W. Freund, H. Zha, A look-ahead strategy for the solution of general Hankel systems, Numer. Math. 64 (1993) 295–322.
[26] K. Glover, All optimal Hankel norm appoximations of linear multivariable systems and their L_∞ error bounds, Int. J. Control 39 (6) (1984) 1115–1193.
[27] I. Gohberg, L.A. Sakhnovich (Eds), Matrix and Operator Valued Functions, Operator Theory: Advances in Applications, Vol. 72, Birkhäuser, Basel, 1994.
[28] W.B. Gragg, Stabilization of the UHQR-algorithm, in: Z. Chen, Y. Li, C.A. Michelli, Y. Xu (Eds.), Advances in Computational Mathematics, Lecture Notes in Pure and Applied Mathematics, Vol. 202, Marcel Dekker, New York, 1999, pp. 139–154.
[29] W.B. Gragg, A. Lindquist, On the partial realization problem, Linear Algebra Appl. 50 (1983) 277–319.

[30] U. Grenander, G. Szegő, Toeplitz Forms and Their Applications, University of California Press, Berkley, 1958.
[31] A. Halanay, V. Ionescu, Time-Varying Discrete Linear Systems, Operator Theory: Advances and Applications, Vol. 68, Birkhäuser, Basel, 1994.
[32] B. Hanzon, The area enclosed by the oriented Nyquist diagram and the Hilbert-Schmidt-Hankel norm of a linear system, IEEE Trans. Automat. Control 37 (6) (1992) 835–839.
[33] E. Hayashi, L.N. Trefethen, M.H. Gutknecht, The CF table, Constr. Approx. 6 (1990) 195–223.
[34] G. Heinig, K. Rost, Algebraic Methods for Toeplitz-like Matrices and Operators, Akademie, Berlin, also Birkhäuser, Basel, 1984.
[35] B.L. Ho, R.E. Kalman, Effective construction of linear state-variable models from input/output functions, Regelungstechnik 14 (12) (1996) 545–548.
[36] I.C.F. Ipsen, Systolic algorithms for the parallel solution of dense symmetric positive definite Toeplitz systems, in: M. Schultz (Ed.), Numerical Algorithms for Modern Parallel Computer Architectures, Springer, New York, 1988, pp. 85–108.
[37] T. Kailath, A.H. Sayed, Displacement structure: theory and applications, SIAM Rev. 37 (1995) 297–386.
[38] M. Kimura, Chain Scattering Approach to H-Infinity-Control, Birkhäuser, Basel, 1997.
[39] L. Ljung, in: System Identification: Theory for the User, 2nd Edition, Information and System Sciences Series, Prentice-Hall, Englewood Cliffs, NJ, 1999.
[40] B.C. Moore, Principal component analysis in linear systems: controllability, observability and model reduction, IEEE Trans. Automat. Control 26 (1981) 17–32.
[41] B. Ninness, F. Gustafsson, A unified construction of orthogonal bases for system identification, IEEE Trans. Automat. Control 42 (1997) 515–522.
[42] J. Partington, Interpolation, Identification and Sampling, London Mathematical Society Monographs, Vol. 17, Oxford University Press, Oxford, 1997.
[43] V.P. Potapov, The Multiplicative Structure of J-Contractive Matrix Functions, American Mathematical Society Translation Series 2, Vol. 15, American Mathematics Society, Providence, RI, 1960, pp. 131–243.
[44] G. Schelfhout, Model reduction for control design, Ph.D. Thesis, Electrotechnical Engineering ESAT, K.U. Leuven, January 1996.
[45] L.N. Trefethen, Near circularity of the error curve in complex Chebyshev approximation, J. Approx. Theory 31 (1981) 344–367.
[46] M. Van Barel, A. Bultheel, Orthogonal polynomial vectors and least squares approximation for a discrete inner product, Electron. Trans. Numer. Anal. 3 (1995) 1–23.
[47] P. Van Overschee, B. De Moor, Subspace Identification for Linear Systems: Theory, Implementation, Applications, Kluwer, Dordrecht, 1996.
[48] N. Wiener, Extrapolation, Interpolation and Smoothing of Time Series, Wiley, New York, 1949.
[49] N. Wiener, P. Masani, The prediction theory of multivariate stochastic processes, I, The regularity condition, Acta Math. 98 (1957) 111–150.