Approximation of linear dynamical systems*

A.C. Antoulas
Department of Electrical and Computer Engineering
Rice University
Houston, Texas 77251-1892

E-mail: acarice.edu - Fax: +1-713 524-5237

March 2, 1998

Abstract

Approximation is an important issue in the theory and practice of dynamical systems. In this essay we will review a theory of approximation of linear dynamical systems which has two important properties. First, it preserves stability and second, explicit a priori bounds for the norm of the error can be derived. Main ingredients of this theory are: the 2-norms used to measure the quantities involved, in particular the Hankel norms, the infinity norms, and a set of invariants called the Hankel singular values. The main tool for the construction of approximants is the all-pass dilation.

Keywords: linear systems, approximation, 2-norms, infinity norm, Hankel operator, Hankel singular values, balancing, balanced truncation, Lyapunov equations, all-pass systems, all-pass dilation.

*This is article **#1049**, prepared for the **Encyclopedia of Electrical and Electronics Engineering**, edited by J.G. Webster, published by John Wiley and Sons, Inc.

1 Introduction

Quite often there is a need for a reduced complexity model of a given high complexity dynamical system. Thus the problem of reducing the complexity of a given system, while retaining key properties, arises. Two important features are

- Appropriate quantization of the approximation error, and
- Retention of stability: if the original high complexity model is stable, the same should hold for the reduced complexity model.

The importance of the first item cannot be overemphasized: in approximating a system, one wishes to have some measure of what has been eliminated.

In the sequel we will deal with linear, time-invariant, discrete- and continuous-time, finite-dimensional, systems described by convolution sums or integrals. In addition, we will only consider the approximation of *stable* systems. Most of the approximation methods available - e.g Padé approximation - fail to satisfy the requirements listed above. We will discuss the theory of *Hankel-norm approximation* and the related *approximation by balanced truncation*. The size of systems - and error systems - is measured in terms of appropriately defined 2-norms.

This approach to model reduction has an interesting history. For operators in finite-dimensional spaces (matrices), the problem of optimal approximation by operators of lower rank is solved by the Schmidt-Mirsky theorem (see [23]). Although this is *not* a convex problem and consequently conventional optimization methods do not apply, it can be solved explicitly by using an ad-hoc tool: the *singular value decomposition* (SVD) of the operator. The solution involves the *truncation* of *small* singular values.

In the same vein, a linear dynamical system can be represented by means of a structured (Hankel) linear operator in appropriate infinite dimensional spaces. The AAK (Adamjan-Arov-Krein) theory (see [1] and [2]), generalizes the Schmidt-Mirsky result to dynamical systems, that is to stuctured operators in infinite-dimensional spaces. This is also known as *optimal approximation in the Hankel norm*. The original setting of the AAK theory was functional analytic. Subsequent developments due to Glover [14], resulted in a simplified linear algebraic framework. This setting made the theory quite transparent by providing explicit formulae for the quantities involved.

The Hankel norm approximation problem can be addressed and solved both for discrete- and continuous-time systems. However, the following is a fact: the discrete-time case is closer to that of finite-dimensional operators and to the Schmidt-Mirsky result. Therefore, the intuitive understanding of the results in this case is more straightforward. In the continuous-time case on the other hand, while the interpretation of the results is less intuitive than for the discrete-time case, the corresponding formulae turn out to be *simpler* (see remark 3.1). In the sequel, because of this dichotomy we will first state the results in the discrete-time case, trying to make connections and draw parallels with the Schmidt-Mirsky theorem (section 2). The numerous formulae for constructing approximants however, will be given for the continuous-time case (section 3).

Hankel-norm approximation as well as model reduction by balanced truncation, inherit an important property from the Schmidt-Mirsky result: unitary operators form the building blocks of this method and cannot be approximated; a multiple singular value indicates that the original operator contains a unitary part. This unitary part has to either be included in the approximant as a whole or discarded as a whole; it cannot be truncated. The same holds for the extension to Hankel-norm approximation. The fundamental building blocks are all-pass systems, and a multiple

Hankel singular value indicates the existence of an all-pass subsystem. In the approximation, this subsystem will either be eliminated or included as a whole; it cannot be truncated. Consequently, it turns out that the basic operation involved in constructing approximants is *all-pass dilation*, i.e. the extention of the number of states of the original system so that the aggregate becomes all-pass (see main theorem (2.3)).

The paper contains three sections. Section 2 reviews 2-norms and induced 2-norms in finite dimensions and states the Schmidt-Mirsky result. The second half of this section presents the appropriate generalization of these concepts for linear discrete-time systems. An operator which is intrinsically attached to a linear system $\mathbf{\Sigma}$ is the *convolution operator* $\mathcal{S}_{\mathbf{\Sigma}}$. Therefore, the problem of optimal approximation in the 2-induced norm of this operator arises naturally. For reasons explained in section 2.5 however, it is presently not possible to solve this problem, except in some special cases (see remark 3.2(b)). A second operator, the *Hankel operator* $\mathcal{H}_{\mathbf{\Sigma}}$, is obtained by restricting the domain and the range of the convolution operator. The 2-induced norm of $\mathcal{H}_{\mathbf{\Sigma}}$ is the *Hankel norm* of $\mathbf{\Sigma}$. It turns out that the optimal approximation problem is solvable in the Hankel norm of $\mathbf{\Sigma}$. This is the AAK theorem, stated in section 2.6.1. The final sub-section 2.6.2, formulates the main result of optimal and sub-optimal approximation in the Hankel norm. The fundamental construction involved is the *all-pass dilation* $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$.

The most natural norm for the approximation problem is the 2-induced norm of the convolution operator, which is equal to the infinity norm of the associated rational transfer function (i.e. the maximum of the amplitude Bode plot). However only a different problem can be solved, namely the approximation in the 2-induced norm of the Hankel operator (which is different from the convolution operator). Although the problem solved is not the same as the original one, the singular values of the Hankel operator turn out to be important invariants. Among other things, they provide *a priori computable bounds* - both upper and lower - for the infinity norm of the error systems.

Section 3 presents the fundamentals of continuous-time linear systems. The convolution and the Hankel operators for continuous-time systems are introduced together with the grammians and the computation of the singular values of the Hankel operator $\mathcal{H}_{\mathbf{\Sigma}}$. These are followed by Lyapunov equations, inertia results, and state-space characterizations of all-pass systems - all important ingredients of the theory. Subsequently various formulae for optimal and sub-optimal approximants are given. These formulae are presented for continuous-time systems since (as mentioned earlier) they are simpler than their discrete-time counterparts. First comes an input/output approach applicable to single-input single-output systems (section 3.2.1); it involves the solution of a polynomial equation which is straightforward to set-up and solve. The remaining formulae are all state-space based. The following cases are treated (in increasing complexity): square systems, sub-optimal case (section 3.2.2); square systems, optimal case (section 3.2.3); general systems, suboptimal case (section 3.2.4). These are followed by the important section 3.2.5, on error bounds; these concern the 2-induced norms of both the Hankel and the convolution operators of error systems. The final section 3.3, presents the closely related approximation method by balanced truncation. Again error bounds are available, although balanced approximants satisfy no optimality properties.

The paper concludes with 3 examples and a brief overview of some selected recent developments, including an overview of the approximation problem for *unstable* systems.

Besides the original sources , namely [1], [2], and [14], parts of the material presented below can also be found in the books [15], [25], and in the lecture notes [5]. For the mathematical background needed we refer to [10].

Contents

List of Figures

2 The Schmidt-Mirsky theorem and the AAK generalization

In this section we will first state the Schmidt-Mirsky theorem concerned with the approximation of finite-dimensional unstructured operators in the 2-norm, which is the operator norm induced by the vector Euclidean norm. Then, discrete-time linear dynamical systems are introduced together with the associated *convolution operator*; this is a structured operator defined on the space of square summable sequences. It is argued in section 2.5 that due to the fact that the convolution operator has a continuous spectrum, it is not clear how the Schmidt-Mirsky result might be generalized in this case. However, roughly speaking by restricting the domain and the range of the convolution operator, the *Hankel operator* is obtained. It turns out that the Schmidt-Mirsky result can be generalized in this case; this is the famous AAK theorem given in section 2.6.1. The main result in finally formulated in section 2.6.2.

2.1 2-norms and induced 2-norms in finite dimensions

The Euclidean or 2-norm of $x \in \mathbb{R}^n$ is defined as

$$\|x\|_2 := \sqrt{x_1^2 + \cdots + x_n^2}$$

Given the linear map $A: \mathbb{R}^n \to \mathbb{R}^m$, the norm induced by the Euclidean norm in the domain and range of A, is the *2-induced* norm:

$$\|A\|_{2-\text{ind}} := \sup_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2} \tag{2.1}$$

It readily follows that

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{x^* A^* A x}{x^* x} \leq \lambda_{\max}(A^* A)$$

where $(\cdot)^*$ denotes complex conjugation and transposition. By choosing x to be the eigenvector corresponding to the largest eigenvalue of A^*A, the above upper bound is attained, and hence the 2-induced norm of A is equal to the square root of the largest eigenvalue of A^*A:

$$\|A\|_{2-\text{ind}} = \sqrt{\lambda_{max}(A^* A)} \tag{2.2}$$

The $m \times n$ matrix A can also be considered as an element of the $(m+n)$-dimensional space $\mathbb{R}^{m+n}$. The Euclidean norm of A in this space is called the *Frobenius norm*:

$$\|A\|_F := \sqrt{\text{trace}(A^* A)} = \sqrt{\sum_{i,j} A_{i,j}^2} \tag{2.3}$$

The Frobenius norm is not an induced norm. It satisfies: $\|A\|_{2-\text{ind}} \leq \|A\|_F$.

2.2 The SVD and the Schmidt-Mirsky theorem

Consider a rectangular matrix $A \in \mathbb{R}^{n \times m}$; let the eigenvalue decomposition of the symmetric matrices A^*A and AA^* be:

$$A^*A = VS_VV^*, \;\; AA^* = US_UU^*$$

where U, V are (square) orthogonal matrices of size n, m respectively (i.e. $UU^* = I_n$, $VV^* = I_m$). Furthermore S_V, S_U are diagonal, and assuming that $n \leq m$:

$$S_V = \text{diag}\,(\sigma_1^2, \cdots, \sigma_n^2, 0, \cdots, 0), \;\; S_U = \text{diag}\,(\sigma_1^2, \cdots, \sigma_n^2), \;\; \sigma_i \geq \sigma_{i+1} \geq 0$$

Given the orthogonal matrices U, V and the nonnegative real numbers $\sigma_1, \cdots, \sigma_n$, we have

$$A = USV^*$$

where S is a matrix of size $n \times m$ with σ_i on the diagonal and zeros elsewhere. This is the *singular value decomposition (SVD)* of A; σ_i, $i = 1, \cdots, n$, are the *singular values* of A, while the columns u_i, v_i of U, V are the *left*, *right singular vectors* of A, respectively. As a consequence, the *dyadic decomposition* of A follows:

$$A = \sum_{i=1}^{n} \sigma_i u_i v_i^* \tag{2.4}$$

This is a decomposition in terms of rank one matrices; the rank of the sum of any k terms is k. In terms of the singular values the Frobenius norm of A defined by (2.3), is:

$$\|A\|_F = \sqrt{\sigma_1^2 + \cdots + \sigma_n^2}$$

The following central problem can now be addressed.

Problem 2.1 Optimal low-rank approximation. *Given the finite matrix A, find a matrix X of the same size but lower rank such that the 2-induced norm of the error $E := A - X$ is minimized.*

The solution of this problem is provided by the Schmidt-Mirsky theorem (see e.g. [23], page 208):

Theorem 2.1 Schmidt-Mirsky. *Given is the matrix A of rank n. For all matrices X of the same size and rank at most $k < n$, there holds:*

$$\|A - X\|_{2-\text{ind}} \geq \sigma_{k+1}(A) \tag{2.5}$$

The lower bound $\sigma_{k+1}(A)$ of the error, is attained by X_*, which is obtained by truncating the dyadic decomposition of A, to the leading k terms:

$$X_* := \sum_{i=1}^{k} \sigma_i u_i v_i^* \tag{2.6}$$

Remark 2.1 (a) *The importance of this theorem lies in the fact that it establishes a relationship between the rank k of the approximant, and the $(k+1)^{st}$ largest singular value of A.*

(b) *The minimizer X_* given above is not unique, since each member of the family of approximants*

$$X(\eta_1, \cdots, \eta_k) := \sum_{i=1}^{k} (\sigma_i - \eta_i) u_i v_i^*, \;\; 0 \leq \eta_i \leq \sigma_{k+1}, \;\; i = 1, \cdots, k \tag{2.7}$$

attains the lower bound, namely $\sigma_{k+1}(A)$.

(c) *The problem of minimizing the 2-induced norm of $A - X$ over all matrices X of rank at most k, is a non-convex optimization problem, since the rank of the sum (or of a linear combination) of two rank k matrices is, in general, not k. Therefore, there is little hope of solving it using conventional optimization methods.*

(d) *If the error in the above approximation is measured in terms of the Frobenius norm, the lower bound in (2.5) is replaced by $\sqrt{\sigma_{k+1}^2 + \cdots + \sigma_n^2}$. In this case, X_* defined by (2.6) is the unique optimal approximant of A, which has rank k.* ∎

2.3 2-norms and induced 2-norms in infinite dimensions

The 2-norm of the infinite sequence $x : \mathbb{Z} \to \mathbb{R}$, is

$$\|x\|_2 := \sqrt{\cdots + x(-1)^2 + x(0)^2 + x(1)^2 + \cdots}$$

The space of all sequences over $\mathbb{Z}$ which have finite 2-norm is denoted by $\ell_2(\mathbb{Z})$; ℓ_2 is known as the *Lebesgue space* of square-summable sequences. Similarly, for x defined over the negative or positive integers $\mathbb{Z}_-$, $\mathbb{Z}_+$; the corresponding spaces of sequences having finite 2-norm are denoted by $\ell_2(\mathbb{Z}_-)$, $\ell_2(\mathbb{Z}_+)$. The 2-norm of the matrix sequence $X : \mathbb{Z} \to \mathbb{R}^{p\times m}$, is defined as:

$$\|X\|_2 := \sqrt{\cdots + \|X(-1)\|_F^2 + \|X(0)\|_F^2 + \|X(1)\|_F^2 + \cdots}$$

The space of all $p \times m$ sequences having finite 2-norm is denoted by $\ell_2^{p\times m}(\mathbb{Z})$. Given the linear map $A:\ X \to Y$, where X, Y are subspaces of $\ell_2(\mathbb{Z})$, the norm induced by the 2-norm in the domain and range is the *2-induced* norm, defined by (2.1); again (2.2) holds.

2.4 Linear, discrete-time systems

A linear, time-invariant, finite-dimensional, discrete-time system $\mathbf{\Sigma}$ is defined as follows:

$$\mathbf{\Sigma}:\ \begin{array}{rcl} x(t+1) & = & Ax(t)+Bu(t) \\ y(t) & = & Cx(t)+Du(t) \end{array},\quad t\in\mathbb{Z}$$

$x(t) \in \mathbb{R}^n$ is the value of the *state* at time t, $u(t) \in \mathbb{R}^m$, $y(t) \in \mathbb{R}^p$ the values of the input, output at time t, respectively; A, B, C, D are constant maps. For simplicity, we will use the notation

$$\mathbf{\Sigma} := \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \in \mathbb{R}^{(n+p)\times(n+m)} \tag{2.8}$$

The *dimension* (or *order*) of the system is n: $\dim \mathbf{\Sigma} = n$. We will denote the unit step function by $\mathbb{I}$ ($\mathbb{I}(t) = 1$, for $t > 0$, and zero otherwise) and the Kronecker delta by δ ($\delta(0) = 1$, and zero otherwise). The *impulse response* $h_{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$ is:

$$h_{\mathbf{\Sigma}}(t) = CA^{t-1}B\mathbb{I}(t) + D\delta(t) \tag{2.9}$$

To $\mathbf{\Sigma}$ we associate the *convolution operator* $\mathcal{S}_{\mathbf{\Sigma}}$ which maps inputs u into outputs y:

$$\mathcal{S}_{\mathbf{\Sigma}}:\ u \longmapsto y \quad \text{where} \quad y(t) = (h_{\mathbf{\Sigma}} * u)(t) := \sum_{\tau=-\infty}^{t} h_{\mathbf{\Sigma}}(t-\tau)u(\tau),\ t\in\mathbb{Z} \tag{2.10}$$

This convolution sum can also be written in matrix notation:

$$\left(\begin{array}{c} \vdots \\ y(-2) \\ y(-1) \\ \hline y(0) \\ y(1) \\ \vdots \end{array}\right) = \underbrace{\left(\begin{array}{cccc|ccc} \ddots & & & & & & \\ & \cdots & h(0) & & & & \\ & \cdots & h(1) & h(0) & & & \\ \hline & \cdots & h(2) & h(1) & h(0) & & \\ & \cdots & h(3) & h(2) & h(1) & h(0) & \\ & & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right)}_{\mathcal{S}_{\mathbf{\Sigma}}} \left(\begin{array}{c} \vdots \\ u(-2) \\ u(-1) \\ \hline u(0) \\ u(1) \\ \vdots \end{array}\right) \tag{2.11}$$

$\mathcal{S}_{\mathbf{\Sigma}}$ has (block) *Toeplitz* and lower-triangular structure. The rational $p \times m$ matrix function

$$H_{\mathbf{\Sigma}}(z) := \sum_{t=0}^{\infty} h_{\mathbf{\Sigma}}(t) z^{-t} = C(zI - A)^{-1}B + D = \left[\frac{p_{ij}(z)}{q_{ij}(z)}\right],\ 1 \le i \le p,\ 1 \le j \le m$$

is called the *transfer function* of $\mathbf{\Sigma}$. The system is *stable* if the eigenvalues of A are inside the unit disk in the complex plane, i.e. $|\lambda_i(A)| < 1$; This condition is equivalent to the impulse response being an ℓ_2 matrix sequence: $h_{\mathbf{\Sigma}} \in \ell_2^{p\times m}(\mathbb{Z})$. If $\mathbf{\Sigma}$ is not stable, its impulse response does not have a finite 2-norm. However, if A has no eigenvalues on the unit circle, the impulse response can be re-interpreted so that it *does* have a finite 2-norm. This is done next. Let T be a basis change in the state space such that the matrices A, B, C can be partitioned as:

$$A = \begin{pmatrix} A_+ & 0 \\ 0 & A_- \end{pmatrix},\ B = \begin{pmatrix} B_+ \\ B_- \end{pmatrix},\ C = (C_+ \ \ C_-)$$

where the eigenvalues of A_+, A_- are inside, outside of the unit disk respectively. In contrast to (2.9), the ℓ_2-impulse response denoted by $h_{\mathbf{\Sigma}2}$ of $\mathbf{\Sigma}$, is defined as follows:

$$\begin{aligned} h_{\mathbf{\Sigma}2} &:= h_{2+} + h_{2-} \ \ \text{where} \\ h_{2+}(t) &:= C_+ A_+^t B_+ \mathbb{I}(t) + \delta(t) D \\ h_{2-}(t) &:= C_- A_-^t B_- \mathbb{I}(-t) \end{aligned} \tag{2.12}$$

where as before $\mathbb{I}$ is the unit step function and δ the Kronecker symbol. Accordingly, we will write $\mathbf{\Sigma}_2 = \mathbf{\Sigma}_+ + \mathbf{\Sigma}_-$. Notice that the algebraic expression for the transfer function remains the same in both cases:

$$H_{\mathbf{\Sigma}2}(z) = H_{2+}(z) + H_{2-}(z) = C_+(zI - A_+)^{-1}B_+ + D + C_-(zI - A_-)^{-1}B_- = H_{\mathbf{\Sigma}}(z)$$

What is modified is the region of convergence of $H_{\mathbf{\Sigma}}$ from

$$|\lambda_{max}(A_-)| < |z| \qquad \text{to} \qquad |\lambda_{max}(A_+)| < |z| < |\lambda_{min}(A_-)|$$

This is equivalent to trading the lack of stability (poles outside the unit disk) for the lack of causality ($h_{\mathbf{\Sigma}2}$ is non-zero for negative time). Thus a system with poles both inside and outside of the unit disk (but not *on* the unit circle) will be interpreted as a possible anti-stable ℓ_2 system $\mathbf{\Sigma}_2$, by defining the impulse response to be non-zero for negative time. Consequently $h_{\mathbf{\Sigma}2} \in \ell_2(\mathbb{Z})$. The matrix representation of the corresponding convolution operator is:

$$\mathcal{S}_{\mathbf{\Sigma}2} = \left(\begin{array}{ccc|ccc} \ddots & \vdots & \vdots & \vdots & \vdots & \\ \cdots & h_{2+}(0) & h_{2-}(-1) & h_{2-}(-2) & h_{2-}(-3) & \cdots \\ \cdots & h_{2+}(1) & h_{2+}(0) & h_{2-}(-1) & h_{2-}(-2) & \cdots \\ \hline \cdots & h_{2+}(2) & h_{2+}(1) & h_{2+}(0) & h_{2-}(-1) & \cdots \\ \cdots & h_{2+}(3) & h_{2+}(2) & h_{2+}(1) & h_{2+}(0) & \cdots \\ & \vdots & \vdots & \vdots & \vdots & \ddots \end{array}\right) \tag{2.13}$$

Notice that $\mathcal{S}_{\mathbf{\Sigma}2}$ has block Toeplitz structure, but is no longer lower triangular. In the sequel: **all discrete-time systems $\mathbf{\Sigma}$ with no poles on the unit circle, will be interpreted as ℓ_2-systems $\mathbf{\Sigma}_2$. For simplicity of notation however, they will still be denoted by $\mathbf{\Sigma}$.**

They are composed of two subsystems: the stable and causal part Σ_+, and the stable but anti-causal part Σ_-: $\Sigma = \Sigma_+ + \Sigma_-$.

For *square* ℓ_2-systems, i.e. ℓ_2-systems having the same number of inputs and outpus $p = m$, the class of *all-pass* ℓ_2-systems is defined as follows: for all pairs u, y satisfying (2.10), there holds

$$\|y\|_2 = \alpha\|u\|_2 \tag{2.14}$$

for some fixed positive constant α. This is equivalent to:

$$\mathcal{S}_{\mathbf{\Sigma}}^*\mathcal{S}_{\mathbf{\Sigma}} = \alpha^2 I \quad \Leftrightarrow \quad H_{\mathbf{\Sigma}}^*(z^{-1})H_{\mathbf{\Sigma}}(z) = \alpha^2 I_m, \ |z| = 1$$

This last condition says that the transfer function (scaled by α) is a unitary matrix on the unit circle.

2.5 Approximation of Σ in the 2-induced norm of the convolution operator

Let $\mathbf{\Sigma}$ be an ℓ_2 system. The convolution operator $\mathcal{S}_{\mathbf{\Sigma}}$ can be considered as a map:

$$\mathcal{S}_{\mathbf{\Sigma}}: \ \ell_2^m(\mathbb{Z}) \longrightarrow \ell_2^p(\mathbb{Z})$$

The 2-induced norm of $\mathbf{\Sigma}$ is defined as the 2-induced norm of $\mathcal{S}_{\mathbf{\Sigma}}$:

$$\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} := \|\mathcal{S}_{\mathbf{\Sigma}}\|_{2-\mathrm{ind}} = \sup_{u \neq 0} \frac{\|\mathcal{S}_{\mathbf{\Sigma}}u\|_2}{\|u\|_2}$$

Due to the equivalence between the time and frequency domains (the Fourier transform is an isometric isomorphism), this norm can also be defined in the frequency domain. In particular

$$\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} = \sup_{|z|=1} \sigma_{max} H_{\mathbf{\Sigma}}(z) =: \|H_{\mathbf{\Sigma}}\|_{\ell_\infty}$$

This latter quantity is known as the ℓ_∞-norm of $H_{\mathbf{\Sigma}}$. If the system is single-input sinlge-output, it is the supremum of the amplitude Bode plot. In case that $\mathbf{\Sigma}$ is stable, $H_{\mathbf{\Sigma}}$ is analytic outside the unit disk, and there holds:

$$\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} = \sup_{|z|\geq 1} \sigma_{max} H_{\mathbf{\Sigma}}(z) =: \|H_{\mathbf{\Sigma}}\|_{h_\infty}$$

This is known as the h_∞ norm of H. For simplicity, we will use the notation

$$\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} := \|\mathcal{S}_{\mathbf{\Sigma}}\|_{2-\mathrm{ind}} = \|H_{\mathbf{\Sigma}}\|_\infty \tag{2.15}$$

It will be clear from the context whether the subscript ∞ stands for the ℓ_∞ norm or the h_∞ norm. If $\mathbf{\Sigma}$ is all-pass, i.e. (2.14) is satisfied, then $\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} = \|H_{\mathbf{\Sigma}}\|_\infty = \alpha$.

Our aim is the generalization of the Schmidt-Mirsky result for an appropriately defined operator. It is most natural to explore the possibility of optimal approximation of $\mathbf{\Sigma}$ in the 2-induced norm of the convolution operator $\mathcal{S}_{\mathbf{\Sigma}}$. In this regard the following result holds.

Proposition 2.1 *Let $\sigma_* := \inf_{|z|=1} \sigma_{min}(H_{\mathbf{\Sigma}}(z))$, while $\sigma^* := \sup_{|z|=1} \sigma_{max}(H_{\mathbf{\Sigma}}(z))$. Every σ in the interval $[\sigma_*, \sigma^*]$, is a singular value of the operator $\mathcal{S}_{\mathbf{\Sigma}}$.*

We conclude that since the singular values of $\mathcal{S}_{\mathbf{\Sigma}}$ form a continuum, it is not a priori clear how the Schmidt-Mirsky result might be generalized to the convolution operator of $\mathbf{\Sigma}$.

Remark 2.2 *Despite the above conclusion, the approach discussed below yields as byproduct the solution of this problem in the special case of the one-step model reduction; see remark 3.2*(**b**).
■

2.6 Approximation of Σ in the 2-induced norm of the Hankel operator

The next attempt to address this approximation problem is by defining a *different* operator attached to the system $\mathbf{\Sigma}$. Recall that $\ell_2^m(\mathcal{I})$, denotes the space of square-summable sequences of vectors, defined on the interval $\mathcal{I}$, with entries in $\mathbb{R}^m$. Given the *stable* and *causal* system $\mathbf{\Sigma}$, the following operator is defined by restricting the domain and the range of the convolution operator (2.10):

$$\begin{aligned} \mathcal{H}_{\mathbf{\Sigma}} : \ell^m(\mathbb{Z}_-) &\longrightarrow \ell^p(\mathbb{Z}_+) \\ u_- &\longmapsto y_+ \;\text{ where }\; y_+(t) = \sum_{\tau=-\infty}^{-1} h_{\mathbf{\Sigma}}(t-\tau)u_-(\tau),\; t \geq 0 \end{aligned} \tag{2.16}$$

$\mathcal{H}_{\mathbf{\Sigma}}$ is called the *Hankel operator* of $\mathbf{\Sigma}$. Its matrix representation in the canonical bases is:

$$\begin{pmatrix} y(0) \\ y(1) \\ y(2) \\ \vdots \end{pmatrix} = \underbrace{\begin{pmatrix} h(1) & h(2) & h(3) & \cdots \\ h(2) & h(3) & h(4) & \cdots \\ h(3) & h(4) & h(5) & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix}}_{\mathcal{H}_{\mathbf{\Sigma}}} \begin{pmatrix} u(-1) \\ u(-2) \\ u(-3) \\ \vdots \end{pmatrix} \tag{2.17}$$

Thus the Hankel operator of $\mathbf{\Sigma}$ maps *past* inputs into *future* outputs. It has a number of properties given next. The first one for single-input single-output systems is due to Kronecker.

Proposition 2.2 *Given the system $\mathbf{\Sigma}$ defined by (2.8), the rank of $\mathcal{H}_{\mathbf{\Sigma}}$ is at most n. The rank is exactly n if, and only if, the system $\mathbf{\Sigma}$ is reachable and observable. Furthermore, if $\mathbf{\Sigma}$ is stable, $\mathcal{H}_{\mathbf{\Sigma}}$ has a finite set of non-zero singular values.*

In order to compute the singular values of the Hankel operator, we define the *reachability* matrix $\mathcal{R}$ and the *observability* matrix $\mathcal{O}$:

$$\mathcal{R}(A,B) = [B \;\; AB \;\; A^2B \;\; \cdots\,], \quad \mathcal{O}(C,A) = [\mathcal{R}(A^*,C^*)]^*$$

both of these matrices have rank at most n. The *reachability, observability grammians* are:

$$\mathcal{P} := \mathcal{R}\mathcal{R}^* = \sum_{t\geq 0} A^t BB^*(A^*)^t, \quad \mathcal{Q} := \mathcal{O}^*\mathcal{O} = \sum_{t \geq 0}(A^*)^t C^*CA^t \tag{2.18}$$

The quantities $\mathcal{P}$, $\mathcal{Q}$ are $n \times n$ symmetric, positive semi-definite matrices. They are positive definite if and only if $\mathbf{\Sigma}$ is reachable and observable. The grammians are (the unique) solutions of the linear matrix equations:

$$A\mathcal{P}A^* + BB^* = \mathcal{P}, \;\; A^*\mathcal{Q}A + C^*C = \mathcal{Q} \tag{2.19}$$

which are called *discrete-time Lyapunov* or *Stein* equations.

The (non-zero) singular values σ_i of $\mathcal{H}_{\mathbf{\Sigma}}$ are the square roots of the (non-zero) eigenvalues λ_i of $\mathcal{H}_{\mathbf{\Sigma}}^*\mathcal{H}_{\mathbf{\Sigma}}$. The key to this computation is the fact that the Hankel operator can be factored in the product of the observability and the reachability matrices:

$$\mathcal{H}_{\mathbf{\Sigma}} = \mathcal{O}(C,A)\mathcal{R}(A,B)$$

For $u_i \in \ell_2^m(\mathbb{Z}_-)$, $u_i \neq 0$, there holds:

$$\mathcal{H}_{\mathbf{\Sigma}}^* \mathcal{H}_{\mathbf{\Sigma}} u_i = \sigma_i^2 u_i \quad \Leftrightarrow \quad \mathcal{R}^* \mathcal{O}^* \mathcal{O} \mathcal{R} u_i = \sigma_i^2 u_i \quad \Leftrightarrow \quad \underbrace{\mathcal{R}\mathcal{R}^*}_{\mathcal{P}} \underbrace{\mathcal{O}^* \mathcal{O}}_{\mathcal{Q}} \mathcal{R} u_i = \sigma_i^2 \mathcal{R} u_i$$

Thus, u_i is an eigenfunction of $\mathcal{H}_{\mathbf{\Sigma}}^* \mathcal{H}_{\mathbf{\Sigma}}$ corresponding to the non-zero eigenvalue σ_i^2, iff $\mathcal{R} u_i \neq 0$, is an eigenfunction of the product of the grammians $\mathcal{P}\mathcal{Q}$:

$$\sigma_i^2(\mathcal{H}_{\mathbf{\Sigma}}) = \lambda_i(\mathcal{H}_{\mathbf{\Sigma}}^* \mathcal{H}_{\mathbf{\Sigma}}) = \lambda_i(\mathcal{P}\mathcal{Q}) \tag{2.20}$$

Proposition 2.3 *The non-zero singular values of the Hankel operator $\mathcal{H}_{\mathbf{\Sigma}}$ associated with the stable system $\mathbf{\Sigma}$, are the square roots of the eigenvalues of the product of the grammians $\mathcal{P}\mathcal{Q}$.*

Definition 2.1 *The* **Hankel singular values** *of the stable system $\mathbf{\Sigma}$ as in (2.10), denoted by:*

$$\sigma_1(\mathbf{\Sigma}) > \cdots > \sigma_q(\mathbf{\Sigma}) \text{ with multiplicity } r_i,\ i = 1, \cdots, q,\ \sum_{i=1}^{q} r_i = n \tag{2.21}$$

are the singular values of $\mathcal{H}_{\mathbf{\Sigma}}$ defined by (2.16). The **Hankel norm** *of $\mathbf{\Sigma}$ is the largest Hankel singular value:*

$$\|\mathbf{\Sigma}\|_H := \sigma_1(\mathbf{\Sigma})$$

The Hankel operator of a not necessarily stable ℓ_2 system $\mathbf{\Sigma}$, is defined as the Hankel operator of its stable and causal part $\mathbf{\Sigma}_+$: $\mathcal{H}_{\mathbf{\Sigma}} := \mathcal{H}_{\mathbf{\Sigma}_+}$.

Thus, the Hankel norm of a system having poles both inside and outside the unit circle, is defined to be the Hankel norm of its causal and stable part. In general

$$\|\mathbf{\Sigma}\|_{2-\text{ind}} \geq \|\mathbf{\Sigma}\|_H \tag{2.22}$$

An important property of the Hankel singular values of $\mathbf{\Sigma}$ is that twice their sum provides an upper bound for the 2-induced norm of $\mathbf{\Sigma}$ (see also section 3.2.5).

Lemma 2.1 *Given the stable system $\mathbf{\Sigma}$, with Hankel singular values σ_i, $i = 1, \cdots, n$ (muliplicities included), the following holds true: $\|\mathbf{\Sigma}\|_{2-\text{ind}} \leq 2(\sigma_1 + \cdots + \sigma_n)$.*

2.6.1 The AAK theorem

Consider the stable systems $\mathbf{\Sigma}$, $\mathbf{\Sigma}'$ of dimension n, k, respectively. By proposition 2.2, $\mathcal{H}_{\mathbf{\Sigma}}$ has rank n and $\mathcal{H}_{\mathbf{\Sigma}'}$ has rank k. Therefore, the Schmidt-Mirsky theorem implies that

$$\|\mathcal{H}_{\mathbf{\Sigma}} - \mathcal{H}_{\mathbf{\Sigma}'}\|_{2-\text{ind}} \geq \sigma_{k+1}(\mathcal{H}_{\mathbf{\Sigma}}) \tag{2.23}$$

The question which arises is to find the infimum of the above norm, given the fact that the approximant is structured (block Hankel matrix): $\inf_{\mathbf{\Sigma}'} \|\mathcal{H}_{\mathbf{\Sigma}} - \mathcal{H}_{\mathbf{\Sigma}'}\|_{2-\text{ind}}$. A remarkable result due to Adamjan, Arov and Krein, code-named *AAK result*, asserts that this lower bound is indeed attained for some $\mathbf{\Sigma}'$ or dimension k. The original sources for this result are [1] and [2].

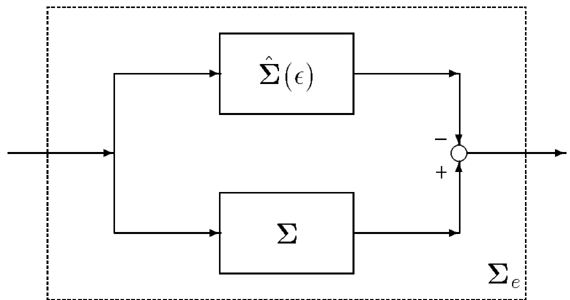


Figure 1: Construction of approximants

Theorem 2.2 AAK theorem. *Given the $\ell_2^{p\times m}(\mathbb{Z}_+)$ sequence of matrices $h = (h(t))_{t>0}$, such that the associated Hankel matrix $\mathcal{H}$ has finite rank n, there exists an $\ell_2^{p\times m}(\mathbb{Z}_+)$ sequence of matrices $h_* = (h_*(t))_{t>0}$, such that the associated Hankel matrix $\mathcal{H}_*$ has rank k and in addition*

$$\|\mathcal{H} - \mathcal{H}_*\|_{2-\mathrm{ind}} = \sigma_{k+1}(\mathcal{H}) \tag{2.24}$$

If p=m=1, the optimal approximant is unique.

This result says that every stable and causal system Σ can be optimally approximated by a stable and causal system Σ_* of lower dimension; the optimality is with respect to the 2-induced norm of the associated Hankel operator.

2.6.2 The main result

In this section we will present the main result. As it turns out one can consider both sub-optimal and optimal approximants within the same framework. Actually, as shown in sections 3.2.2 and 3.2.3, the formulae for sub-optimal approximants are simpler than their optimal counterparts.

Problem 2.2 *Given a stable system Σ, we seek approximants Σ_* satisfying*

$$\sigma_{k+1}(\Sigma) \leq \|\Sigma - \Sigma_*\|_H \leq \epsilon < \sigma_k(\Sigma)$$

This is a genalization of problem 2.1, as well as the problem solved by the AAK theorem. The concept introduced in the next definition is the key to its solution.

Definition 2.2 *Let Σ_e be the parallel connection of Σ and $\hat{\Sigma}$: $\Sigma_e := \Sigma - \hat{\Sigma}$. If Σ_e is an all-pass system with norm ϵ, $\hat{\Sigma}$ is called an ϵ-all-pass dilation of Σ.*

As a consequence of the inertia result of section 3.1.4, the all-pass dilation system has the following crucial property.

Main Lemma 2.1 *Let $\hat{\Sigma}$ be an ϵ-all-pass dilation of Σ, where ϵ satisfies* (2.25). *It follows that $\hat{\Sigma}$ has exactly k poles inside the unit disc, i.e.* $\dim \hat{\Sigma}_+ = k$.

We also restate the analog of the Schmidt-Mirsky result (2.23), applied to dynamical systems:

Proposition 2.4 *Given the stable system Σ, let Σ' have at least k poles inside the unit disc. Then*

$$\|\Sigma - \Sigma'\|_H \geq \sigma_{k+1}(\Sigma)$$

This means that the 2-induced norm of the Hankel operator of the difference between Σ and Σ' is no less than the $(k+1)$st singular value of the Hankel operator of Σ. Finally, recall that if a system has both stable and unstable poles, its Hankel norm, is that of its stable part. We are now ready for the main result.

Theorem 2.3 *Let $\hat{\Sigma}$ be an ϵ-all-pass dilation of the linear, stable, discrete- or continuous-time system Σ, where*

$$\sigma_{k+1}(\Sigma) \leq \epsilon < \sigma_k(\Sigma) \tag{2.25}$$

It follows that $\hat{\Sigma}_+$ has exactly k stable poles and consequently

$$\sigma_{k+1}(\Sigma) \leq \|\Sigma - \hat{\Sigma}\|_H < \epsilon \tag{2.26}$$

In case $\sigma_{k+1}(\Sigma) = \epsilon$,

$$\sigma_{k+1}(\Sigma) = \|\Sigma - \hat{\Sigma}\|_H$$

Proof. The result is a consequence of the following sequence of equalities and inequalities:

$$\sigma_{k+1}(\Sigma) \leq \|\Sigma - \hat{\Sigma}_+\|_H = \|\Sigma - \hat{\Sigma}\|_H \leq \|\Sigma - \hat{\Sigma}\|_\infty = \epsilon$$

The first inequality on the left side is a consequence of the main lemma 2.1, the equality follows by definition, the second inequality follows from (2.22), and the last equality holds by construction, since $\Sigma - \hat{\Sigma}$ is ϵ-all-pass. ■

Remark 2.3 **(a)** *For $\epsilon = \sigma_1(\Sigma)$, the above theorem yields the solution of the Nehari problem, namely to find the best anti-stable approximant of Σ in the 2-induced norm of the convolution operator (i.e. the ℓ_∞-norm).*

(b) *We are given a stable system Σ and seek to compute an approximant in the same class (i.e. stable). In order to achieve this, the construction given above takes us outside this class of systems, since the all-pass dilation system $\hat{\Sigma}$ has poles both inside and outside the unit circle. In terms of matrices, we start with a system whose convolution operator $\mathcal{S}_{\Sigma}$ is a (block) lower triangular Toeplitz matrix. We then compute a (block) Toeplitz matrix $\mathcal{S}_{\hat{\Sigma}}$, which is no longer lower triangular, such that the difference is unitary. It then follows that the lower left-hand portion of $\mathcal{S}_{\hat{\Sigma}}$, which is the Hankel matrix $\mathcal{H}_{\hat{\Sigma}}$ has rank r and approximates the Hankel matrix $\mathcal{H}_{\Sigma}$, so that the 2-norm of the error satisfies (2.25).*

(c) *The sub-optimal and optimal approximants can be constructed using explicit formulae. For continuous-time systems, see section 3.2.* ■

3 Construction of approximants

The purpose of this section is to present and to a certain extent derive, formulae for sub-optimal and optimal approximants in the Hankel norm. **Because of main theorem 2.3, all we need**

is the ability to construct all-pass dilations of a given system. To this goal, the first subsection is dedicated to the presentation of important aspects of the theory of linear, continuous-time systems; these facts are used in the second subsection. The closely related approach to system approximation by balanced truncation is briefly discussed in section 3.3.

3.1 Linear, continuous-time systems

3.1.1 $\mathcal{L}_2$ linear systems

For continuous-time functions, let

$$\mathcal{L}^n(\mathcal{I}) := \{f : \mathcal{I} \longrightarrow \mathbb{R}^n,\ \mathcal{I} \subset \mathbb{R}\}$$

Frequent choices of $\mathcal{I}$: $\mathcal{I} = \mathbb{R}$, $\mathcal{I} = \mathbb{R}_+$ or $\mathcal{I} = \mathbb{R}_-$. The 2-norm of a function f is:

$$\|f\|_2 := \left(\int_{t\in\mathcal{I}} \| f(t) \|_2^2 \, dt\right)^{\frac{1}{2}},\ f \in \mathcal{L}^n(\mathcal{I})$$

The corresponding $\mathcal{L}_2$ space of square integrable functions is:

$$\mathcal{L}_2^n(\mathcal{I}) \ := \ \{f \in \mathcal{L}^n(\mathcal{I}),\ \| f \|_2 < \infty\}$$

The 2-norm of the matrix function $F : \mathcal{I} \to \mathbb{R}^{p\times m}$, is defined as:

$$\|F\|_2 := \left(\int_{t\in\mathcal{I}} \| F(t) \|_F^2 \, dt\right)^{\frac{1}{2}}$$

The space of all $p\times m$ matrix functions having finite 2-norm is denoted by $\mathcal{L}_2^{p\times m}(\mathcal{I})$. Let $A :\ X \to Y$, where X, Y are subspaces of $\mathcal{L}_2(\mathbb{R})$. The 2-induced norm of A is defined as in (2.1), and (2.2) holds true.

In the sequel, we will consider linear, finite-dimensional, time-invariant, continuous-time systems described by the following set of differential and algebraic equations:

$$\mathbf{\Sigma} : \quad \begin{array}{rcl} \frac{dx(t)}{dt} & = & Ax(t) + Bu(t) \\ y(t) & = & Cx(t) + Du(t) \end{array}, \quad t \in \mathbb{R}$$

where $u(t) \in \mathbb{R}^m$, $x(t) \in \mathbb{R}^n$, $y(t) \in \mathbb{R}^p$, are the values of the input, state, output at time t, respectively. The system will be abbreviated as

$$\mathbf{\Sigma} = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) \in \mathbb{R}^{(n+p)\times(n+m)} \tag{3.1}$$

In analogy to the discrete-time case, we define the continuous-time unit step function $\mathbb{I}(t)$ ($\mathbb{I}(t) = 1$ for $t \geq 0$, and zero otherwise) and the delta distribution δ. The impulse response of this system is

$$h_{\mathbf{\Sigma}}(t) = Ce^{At}B\mathbb{I}(t) + \delta(t)D, \tag{3.2}$$

while the transfer function is

$$H_{\mathbf{\Sigma}}(s) = C(sI - A)^{-1}B + D$$

Unless A has all its eigenvalues in the left-half-plane (LHP), the impulse response $h_{\mathbf{\Sigma}}$ is not square integrable, i.e. does not belong to the space $\mathcal{L}_2^{p\times m}(\mathbb{R})$. In this case we shall say that the system is not an $\mathcal{L}_2$-system. If however, A has no eigenvalues on the $j\omega$-axis, it can be interpreted as an $\mathcal{L}_2$-system by appropriate re-definition of the impulse response. As in the discrete-time case, let there be a state transformation such that

$$A = \begin{pmatrix} A_+ & 0 \\ 0 & A_- \end{pmatrix},\ B = \begin{pmatrix} B_+ \\ B_- \end{pmatrix},\ C = (C_+ \ \ C_-)$$

where the eigenvalues of A_+, A_- are in the LHP, RHP (right-half plane) respectively. The $\mathcal{L}_2$-impulse response is defined as follows:

$$\begin{aligned} h_{\mathbf{\Sigma}2} &:= h_{2+} + h_{2-} \quad \text{where} \\ h_{2+}(t) &:= C_+ e^{A_+ t} B_+ \mathbb{I}(t) + \delta(t) D \\ h_{2-}(t) &:= C_- e^{A_- t} B_- \mathbb{I}(-t) \end{aligned} \tag{3.3}$$

As in the discrete-time case

$$H_{\mathbf{\Sigma}2}(s) = H_{2+}(s) + H_{2-}(s) = C_+(sI - A_+)^{-1}B_+ + D + C_-(sI - A_-)^{-1}B_- = H_{\mathbf{\Sigma}}(s)$$

This is equivalent to trading the lack of stability (poles in the RHP) to the lack of causality ($h_{\mathbf{\Sigma}2}$ is non-zero for negative time). What is modified is the region of convergence from

$$\mathcal{R}e(\lambda_{max}(A_-)) < \mathcal{R}e(s) \qquad \text{to} \qquad \mathcal{R}e(\lambda_{max}(A_+)) < \mathcal{R}e(s) < \mathcal{R}(\lambda_{min}(A_-))$$

From now on, **all continuous-time systems $\mathbf{\Sigma}$ with no poles on the imaginary axis, will be interpreted as $\mathcal{L}_2$ systems; to keep the notation simple they will be denoted by $\mathbf{\Sigma}$ instead of $\mathbf{\Sigma}_2$. The stable and causal subsystem will be denoted by $\mathbf{\Sigma}_+$ and the stable but anti-causal one by $\mathbf{\Sigma}_-$: $\mathbf{\Sigma} = \mathbf{\Sigma}_+ + \mathbf{\Sigma}_-$.**

3.1.2 The convolution operator

The convolution operator associated to $\mathbf{\Sigma}$ is defined as follows

$$\mathcal{S}_{\mathbf{\Sigma}} :\ u \longmapsto y \ \text{ where } \ y(t) = (h_{\mathbf{\Sigma}} * u)(t) := \int_{-\infty}^{\infty} h_{\mathbf{\Sigma}}(t-\tau)u(\tau)d\tau,\ t \in \mathbb{R} \tag{3.4}$$

Corresponding to proposition 2.1 we have

Proposition 3.1 *Let $\sigma_* := \inf_{\omega\in\mathbb{R}} \sigma_{min}(H_{\mathbf{\Sigma}}(j\omega))$, while $\sigma^* := \sup_{\omega\in\mathbb{R}} \sigma_{max}(H_{\mathbf{\Sigma}}(j\omega))$. Every σ in the interval $[\sigma_*, \sigma^*]$, is a singular value of the operator $\mathcal{S}_{\mathbf{\Sigma}}$.*

Since the singular values of $\mathcal{S}_{\mathbf{\Sigma}}$ form a continuum the same conclusion as in the discrete-time case follows (see also remark 2.2). Again, the 2-induced norm of $\mathbf{\Sigma}$ turns out to be the infinity norm of the transfer function $H_{\mathbf{\Sigma}}$

$$\|\mathbf{\Sigma}\|_{2-\text{ind}} := \|\mathcal{S}_{\mathbf{\Sigma}}\|_{2-\text{ind}} = \|H_{\mathbf{\Sigma}}\|_\infty \tag{3.5}$$

If $\mathbf{\Sigma}$ is all-pass (2.14) then $\|\mathbf{\Sigma}\|_{2-\text{ind}} = \|H_{\mathbf{\Sigma}}\|_\infty = \alpha$.

3.1.3 All-pass $\mathcal{L}_2$-systems

All-pass systems are an sub-class of $\mathcal{L}_2$-systems. As indicated in theorem 2.3, all-pass systems play an important role in Hankel norm approximation. The following characterization is central for the construction of sub-optimal and optimal approximants (see sections 3.2.2, 3.2.3). The proof is by direct computation; for details see [5].

Proposition 3.2 *Given is $\mathbf{\Sigma}$ as in (3.1), with $p = m$. The following statements are equivalent.*

- *$\mathbf{\Sigma}$ is α-all-pass.*
- *For all input-output pairs (u, y) satisfying $y = h_{\mathbf{\Sigma}} * u$, there holds: $\|y\|_2 = \alpha\|u\|_2$.*
- *$H^*_{\mathbf{\Sigma}}(-j\omega)H_{\mathbf{\Sigma}}(j\omega) = \alpha^2 I_m$*
- *There exists $\mathcal{Q} = \mathcal{Q}^* \in \mathbb{R}^{n\times n}$, such that the following equations are satisfied*

$$\left.\begin{aligned} A^*\mathcal{Q} + \mathcal{Q}A + C^*C = 0 \\ \mathcal{Q}B + C^*D = 0 \\ D^*D = \alpha^2 I_m \end{aligned}\right\} \tag{3.6}$$

- *The solutions $\mathcal{P}$, $\mathcal{Q}$ of the Lyapunov equations*

$$A\mathcal{P} + \mathcal{P}A^* + BB^* = 0, \quad A^*\mathcal{Q} + \mathcal{Q}A + C^*C = 0$$

  satisfy $\mathcal{P}\mathcal{Q} = \alpha^2 I_n$, and in addition $D^*D = \alpha^2 I_m$.

3.1.4 The grammians, Lyapunov equations and an inertia result

For stable systems (i.e. $\mathcal{R}e(\lambda_i(A)) < 0$), the following quantities are defined:

$$\mathcal{P} := \int_0^\infty e^{At}BB^*e^{A^*t}dt, \;\; \mathcal{Q} := \int_0^\infty e^{A^*t}C^*Ce^{At}dt \tag{3.7}$$

They are called the *reachability*, *observability grammians* of $\mathbf{\Sigma}$ respectively. By definition $\mathcal{P}$, $\mathcal{Q}$ are positive semi-definite. It can be shown that reachability and observability of $\mathbf{\Sigma}$, are equivalent to the positive definitness of each one of these grammians. As in the discrete-time case, the grammians are the (unique) solutions of the linear matrix equations

$$A\mathcal{P} + \mathcal{P}A^* + BB^* = 0, \quad A^*\mathcal{Q} + \mathcal{Q}A + C^*C = 0 \tag{3.8}$$

which are known as the continuous-time *Lyapunov equations* (cf. 2.19). Such equations have a remarkable property known as *inertia* result: there is a relationship between the number of eigenvalues of A and (say) $\mathcal{P}$ in the LHP, RHP and the imaginary axis. More precisely, the *inertia* of $A \in \mathbb{R}^{n\times n}$ is:

$$\text{in}\,(A) := \{\nu(A), \delta(A), \pi(A)\} \tag{3.9}$$

where $\nu(A)$, $\delta(A)$, $\pi(A)$ is the number of eigenvalues of A in the LHP, on the imaginary axis, RHP respectively.

Proposition 3.3 *Let A and $X = X^*$ satisfy the Lyapunov equation: $AX + XA^* = R$, where $R \geq 0$. If the pair (A, R) is reachable, then the inertia of A is equal to the inertia of X:* $\text{in}\,(A) = \text{in}\,(X)$.

The crucial lemma 2.1 is based on this inertia result.

Remark 3.1 *One of the reasons why continuous-time formulae are simpler that their discrete-time counterparts is that the Lyapunov equations (3.8) are* linear *in A, while the discrete-time Lyapunov equations (2.19) are* quadratic *in A.* ■

3.1.5 The continuous-time Hankel operator

In analogy to the discrete-time case, we define the operator $\mathcal{H}_{\mathbf{\Sigma}}$, of the stable system $\mathbf{\Sigma}$ which maps *past inputs* into *future outputs*:

$$\begin{array}{rcl} \mathcal{H}_{\mathbf{\Sigma}} : \mathcal{L}_2^m(\mathbb{R}_-) & \longrightarrow & \mathcal{L}_2^p(\mathbb{R}_+) \\ u_- & \longmapsto & y_+ \;\; \text{where} \;\; y_+(t) = \displaystyle\int_{-\infty}^{0} h_{\mathbf{\Sigma}}(t-\tau)u_-(\tau)d\tau, \; t \geq 0 \end{array} \tag{3.10}$$

$\mathcal{H}_{\mathbf{\Sigma}}$ is called the *Hankel operator* of $\mathbf{\Sigma}$. Unlike in the discrete-time case however (see (2.17)) $\mathcal{H}_{\mathbf{\Sigma}}$ has *no* matrix representation.

It turns out that just as in (2.20), it can be shown that the non-zero singular values of the continuous-time Hankel operator are the eigenvalues of the product of the two grammians

$$\sigma_i^2(\mathcal{H}_{\mathbf{\Sigma}}) = \lambda_i(\mathcal{H}_{\mathbf{\Sigma}}^*\mathcal{H}_{\mathbf{\Sigma}}) = \lambda_i(\mathcal{P}\mathcal{Q}) \tag{3.11}$$

The Hankel singular values of $\mathcal{H}_{\mathbf{\Sigma}}$ and their multiplicities will be denoted as in (2.21). Finally,

$$\|\mathbf{\Sigma}\|_{\mathrm{H}} = \|\mathbf{\Sigma}_+\|_{\mathrm{H}} \leq \|\mathbf{\Sigma}\|_{2-\mathrm{ind}}$$

Lemma 2.1 can be strengthened in the continuous-time case (see also section 3.2.5).

Lemma 3.1 *Let the* **distinct** *Hankel singular values of the stable system $\mathbf{\Sigma}$ be σ_i, $i = 1, \cdots, q$; there holds: $\|\mathbf{\Sigma}\|_{2-\mathrm{ind}} \leq 2(\sigma_1 + \;\cdots\; + \sigma_q)$.*

3.1.6 A transformation between continuous and discrete time systems

As mentioned in the introduction, the formulae for the construction of approximants will be given for the continuous-time case because they are generally simpler than their discrete-time counterparts. Thus, in order to apply the formulae to discrete-time systems, the system will have to be transformed to continuous-time first, and at the end the continuous-time optimal or sub-optimal approximants obtained will have to be transformed back to discrete-time.

One transformation between continuous- and discrete-time systems is given by the bilinear transformation $z = \frac{1+s}{1-s}$, of the complex plane onto itself. The resulting relationship between the transfer function $H_c(s)$ of a continuous-time system and that of the corresponding discrete-time system $H_d(z)$ is:

$$H_c(s) = H_d\left(\frac{1+s}{1-s}\right)$$

The state space maps

$$\mathbf{\Sigma}_c := \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right), \quad \mathbf{\Sigma}_d := \left(\begin{array}{c|c} F & G \\ \hline H & J \end{array}\right)$$

are related as given in the table below. Furthermore, the proposition that follows states that the Hankel and infinity norms remain unchanged by this transformation.

Continuous-time		**Discrete-time**
$A,\ B,\ C,\ D$	$z = \frac{1+s}{1-s}$	$F = (I+A)(I-A)^{-1}$ $G = \sqrt{2}(I-A)^{-1}B$ $H = \sqrt{2}C(I-A)^{-1}$ $J = D + C(I-A)^{-1}B$
$A = (F+I)^{-1}(F-I)$ $B = \sqrt{2}(F+I)^{-1}G$ $C = \sqrt{2}H(F+I)^{-1}$ $D = J - H(F+I)^{-1}G$	$s = \frac{z-1}{z+1}$	$F,\ G,\ H,\ J$

Proposition 3.4 *Given the stable continuous-time system $\mathbf{\Sigma}_c$ with grammians $\mathcal{P}_c$, $\mathcal{Q}_c$, let $\mathbf{\Sigma}_d$ with grammians $\mathcal{P}_d$, $\mathcal{Q}_d$, be the discrete-time system obtained by means of the transformation given above. It follows that $\mathcal{P}_c = \mathcal{P}_d$ and $\mathcal{Q}_c = \mathcal{Q}_d$. Furthermore, this bilinear transformation also preserves the infinity norms (i.e. the 2-induced norms of the associated convolution operators): $\|\mathbf{\Sigma}_c\|_\infty = \|\mathbf{\Sigma}_d\|_\infty$.*

3.2 Construction formulae for Hankel-norm approximants

We are ready to give some of the formulae for the construction of sub-optimal and optimal Hankel-norm approximants. As mentioned earlier all formulae describe the construction of all-pass dilation systems (see 2.3). We will concentrate on the following cases (in increasing degree of complexity).

- An input-output construction applicable to scalar systems. Both optimal and suboptimal approximants are treated. The advantage of this approach is that the equations can be set-up in a straightforward manner using the numerator and denominator polynomials of the transfer function of the given system (section 3.2.1).

- A state space based construction method for sub-optimal approximants (section 3.2.2) and for optimal approximants (section 3.2.3) of square systems.

- A state space based parametrization of *all* sub-optimal approximants for general (i.e. not necessarily square) systems (section 3.2.4).

- The optimality of the approximants is with respect to the Hankel norm (2-induced norm of the Hankel operator). Section 3.2.5, gives an account of error bounds for the infinity norm of approximants (2-induced norm of the convolution operator).

- The final section 3.3 discusses model reduction by balanced truncation, which uses the same ingredients as the Hankel norm model reduction theory. The approximants have a number of interesting properties, including the existence of error bounds for the infinity norm of the error. However, no optimality holds.

3.2.1 Input-output construction method for scalar systems

Given the polynomials $a = \sum_{i=0}^{\alpha} a_i s^i$, $b = \sum_{i=0}^{\beta} b_i s^i$, $c = \sum_{i=0}^{\gamma} c_i s^i$, satisfying $c(s) = a(s)b(s)$, the coefficients of the product c are a linear combination of those of b:

$$\underline{c} = \mathbb{T}(a)\underline{b}$$

where $\underline{c} := (c_\gamma \;\; c_{\gamma-1} \;\; \cdots \;\; c_1 \;\; c_0)^* \in \mathbb{R}^{\gamma+1}$, $\underline{b} := (b_\beta \;\; \cdots \;\; b_0)^* \in \mathbb{R}^{\beta+1}$, and $\mathbb{T}(a)$ is a Toeplitz matrix with first column $(a_\alpha \;\; \cdots \;\; a_0 \;\; 0 \;\; \cdots \;\; 0)^* \in \mathbb{R}^{\gamma+1}$, and first row $(a_\alpha \;\; 0 \;\; \cdots \;\; 0) \in \mathbb{R}^{1\times(\beta+1)}$. We will also define the sign matrix

$$\mathbb{K} = \operatorname{diag}\,(\;\cdots\;,\; 1,\; -1,\; 1)$$

of appropriate size. Given a polynomial a with real coefficients, the polynomial c^* is defined as $c(s)^* := c(-s)$. This means that

$$\underline{c^*} = \mathbb{K}\underline{c}$$

The basic construction given in theorem 2.3 hinges on the construction of an ϵ-all-pass dilation $\hat{\mathbf{\Sigma}}$ of $\mathbf{\Sigma}$. Let

$$H_{\mathbf{\Sigma}}(s) = \frac{p(s)}{q(s)},\;\; H_{\hat{\mathbf{\Sigma}}}(s) = \frac{\hat{p}(s)}{\hat{q}(s)}$$

We require that the difference $H_{\mathbf{\Sigma}} - H_{\hat{\mathbf{\Sigma}}} = H_{\mathbf{\Sigma}_e}$, be ϵ-all-pass. Therefore, the problem is: given ϵ, and the polynomials p, q, such that $\deg(p) \le \deg(q) := n$, find polynomials $\hat{p}$, $\hat{q}$, of degree at most n such that

$$\frac{p}{q} - \frac{\hat{p}}{\hat{q}} = \epsilon\frac{q^*\hat{q}^*}{q\hat{q}} \;\;\Leftrightarrow\;\; p\hat{q} - q\hat{p} = \epsilon q^*\hat{q}^* \tag{3.12}$$

This polynomial equation can be rewritten as a matrix equation involving the quantities defined above:

$$\mathbb{T}(p)\underline{\hat{q}} - \mathbb{T}(q)\underline{\hat{p}} = \epsilon\mathbb{T}(q^*)\underline{\hat{q}^*} = \epsilon\mathbb{T}(q^*)\mathbb{K}\underline{\hat{q}}$$

Collecting terms we have

$$\left(\mathbb{T}(p) - \epsilon\mathbb{T}(q^*)\mathbb{K},\;\; -\mathbb{T}(q)\right)\begin{pmatrix}\underline{\hat{q}}\\ \underline{\hat{p}}\end{pmatrix} = 0 \tag{3.13}$$

The solution of this set of linear equations provides the coefficients of the ϵ-all-pass dilation system $\hat{\mathbf{\Sigma}}$. Furthermore, this system can be solved both for the suboptimal $\epsilon \neq \sigma_i$, and the optimal $\epsilon = \sigma_i$, cases. We will illustrate the features of this approach by means of a simple example. For an alternative approach along similar lines, see [11].

Example. Let $\mathbf{\Sigma}$ be a second order system, i.e. $n = 2$. If we normalize the coefficient of the highest power of $\hat{q}$, i.e. $\hat{q}_2 = 1$, we obtain the following system of equations:

$$\underbrace{\begin{pmatrix} 0 & 0 & q_2 & 0 & 0 \\ p_2 - \epsilon q_2 & 0 & q_1 & q_2 & 0 \\ p_1 + \epsilon q_1 & p_2 + \epsilon q_2 & q_0 & q_1 & q_2 \\ p_0 - \epsilon q_0 & p_1 - \epsilon q_1 & 0 & q_0 & q_1 \\ 0 & p_0 + \epsilon q_0 & 0 & 0 & q_0 \end{pmatrix}}_{W(\epsilon)} \begin{pmatrix} \hat{q}_1 \\ \hat{q}_0 \\ \hat{p}_2 \\ \hat{p}_1 \\ \hat{p}_0 \end{pmatrix} = -\begin{pmatrix} p_2 + \epsilon q_2 \\ p_1 - \epsilon q_1 \\ p_0 + \epsilon q_0 \\ 0 \\ 0 \end{pmatrix}$$

This can be solved for all ϵ which are not roots of the equation $\det W(\epsilon) = 0$. The latter is a polynomial equation of second degree ; there are thus two values of ϵ, ϵ_1 and ϵ_2, for which the determinant of W is zero. It can be shown that the roots of this determinant are the *eigenvalues* of the Hankel operator $\mathcal{H}_{\mathbf{\Sigma}}$; since in the SISO case $\mathcal{H}_{\mathbf{\Sigma}}$ is self-adjoint (symmetric) the absolute values of ϵ_1, ϵ_2, are the singular values of $\mathcal{H}_{\mathbf{\Sigma}}$. Thus both sub-optimal and optimal approximants can be computed this way. (See also the first example of section 4.) ■

3.2.2 State-space construction for square systems: suboptimal case

In this section we will discuss the construction of approximants in the case where $m = p$. Actually this is no loss of generality because if one has to apply the algorithm to a non-square system, additional rows or columns of zeros can be applied so as to make the system square. For details we refer to [14]. Consider the system $\mathbf{\Sigma}$ as in (3.1), with $\mathcal{R}e(\lambda_i(A)) < 0$ and $m = p$. For simplicity it is assumed that the D-matrix of $\mathbf{\Sigma}$ is zero. Compute the grammians $\mathcal{P}$, $\mathcal{Q}$ by solving the Lyapunov equations (3.8). We are looking for

$$\hat{\mathbf{\Sigma}} = \left(\begin{array}{c|c} \hat{A} & \hat{B} \\ \hline \hat{C} & \hat{D} \end{array}\right)$$

such that the parallel connection of $\mathbf{\Sigma}$ and $\hat{\mathbf{\Sigma}}$ denoted by $\mathbf{\Sigma}_e$, is ϵ-all-pass. First, notice that $\mathbf{\Sigma}_e$ has the following state-space representation:

$$\mathbf{\Sigma}_e = \left(\begin{array}{cc|c} A & & B \\ & \hat{A} & \hat{B} \\ \hline C & -\hat{C} & -\hat{D} \end{array}\right) \tag{3.14}$$

According to the last characterization of proposition 3.2, $\mathbf{\Sigma}_e$ is ϵ-all-pass iff the corresponding grammians and D matrices satisfy

$$\mathcal{P}_e\mathcal{Q}_e = \epsilon^2 I_{2\mathrm{n}}, \quad \hat{D}^*\hat{D} = \epsilon^2 I_m \tag{3.15}$$

This implies that $\frac{\hat{D}}{\epsilon}$ is a unitary matrix of size m. The key to the construction of $\mathcal{P}_e$ and $\mathcal{Q}_e$, is the *unitary dilation* of $\mathcal{P}$ and $\mathcal{Q}$. Consider the simple case where the grammians are scalar $\mathcal{P} = p$, $\mathcal{Q} = q$ and $\epsilon = 1$. In this case we are looking for filling in the ? so that:

$$\begin{pmatrix} p & ? \\ ? & ? \end{pmatrix}\begin{pmatrix} q & ? \\ ? & ? \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

assuming that $pq \neq 1$. It readily follows that one solution is:

$$\begin{pmatrix} p & 1 - pq \\ 1 - pq & -q(1 - pq) \end{pmatrix}\begin{pmatrix} q & 1 \\ 1 & -\frac{p}{1-pq} \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.16}$$

In the general (non-scalar) case, this suggests defining the quantity:

$$\Gamma := \epsilon^2 - \mathcal{P}\mathcal{Q} \tag{3.17}$$

Assuming that ϵ is not equal to any of the eigenvalues of the product $\mathcal{P}\mathcal{Q}$ (i.e. to any of the singular values of the Hankel operator $\mathcal{H}_{\mathbf{\Sigma}}$) Γ is invertible. Keeping in mind that $\mathcal{P}$ and $\mathcal{Q}$ do not commute, the choice of $\mathcal{P}_e$, $\mathcal{Q}_e$ corresponding to (3.16) is

$$\mathcal{P}_e = \begin{pmatrix} \mathcal{P} & \Gamma \\ \Gamma^* & -\mathcal{Q}\Gamma \end{pmatrix}, \quad \mathcal{Q}_e = \begin{pmatrix} \mathcal{Q} & I_n \\ I_n & -\Gamma^{-1}\mathcal{P} \end{pmatrix} \tag{3.18}$$

Once the matrices $\mathcal{P}_e$ and $\mathcal{Q}_e$ for the dilated system $\mathbf{\Sigma}_e$ have been constructed, the next step is to construct the matrices $\hat{A}$, $\hat{B}$, $\hat{C}$ of the dilation system $\hat{\mathbf{\Sigma}}$. According to (3.6) of proposition 3.2, the Lyapunov equation $A_e^*\mathcal{Q}_e + \mathcal{Q}_e A_e + C_e^* C_e = 0$, and the equation $\mathcal{Q}_e B_e + C_e^* D_e = 0$, have to be satisfied. Solving these two equations (see (A.1)) for the unknown quantities, we obtain:

$$\begin{aligned} \hat{A} &= -A^* + C^*(C\mathcal{P} - \hat{D}B^*)\Gamma^{-*} \\ \hat{B} &= -\mathcal{Q}B + C^*\hat{D} \\ \hat{C} &= (C\mathcal{P} - \hat{D}B^*)\Gamma^{-*} \end{aligned} \tag{3.19}$$

There remains to show is that $\hat{A}$ has exactly k stable eigenvalues. From the Lyapunov equation for $\mathcal{Q}_e$ it also follows that $\hat{\mathcal{Q}} = -\Gamma^{-1}\mathcal{P}$. By construction, see (3.17), Γ has k positive and $n-k$ negative eigenvalues; the same holds true for $\hat{\mathcal{Q}}$ (i.e. in $(\hat{\mathcal{Q}}) = \{k, 0, n-k\}$). Furthermore, the pair $\hat{C}, \hat{A}$ is observable, because otherwise A has eigenvalues on the $j\omega$ axis, which is a contradiction to the original assumption that A is stable. Then by proposition 3.3, the inertia of $-\hat{\mathcal{Q}}$ is equal to that of $\hat{A}$, which completes the proof.

Corollary 3.1 *The system $\hat{\mathbf{\Sigma}}$ has dimension n and the dilated system $\mathbf{\Sigma}_e$ has dimension $2n$. The stable subsystem $\hat{\mathbf{\Sigma}}_+$ has dimension k.*

3.2.3 State-space construction for square systems: optimal case

The construction of the previous section can be extended to include the optimal case. This section presents the formulae which first appeared in section 6 of [14].

In this case we need to construct the all-pass dilation $\mathbf{\Sigma}_e$ for ϵ equal to the lower bound in (2.25), namely $\sigma_{k+1}(\mathbf{\Sigma})$. Thus, σ_{k+1} is an eigenvalue of the product of the grammians $\mathcal{PQ}$, of multiplicity, say r. There exists a basis change in the state space such that

$$\mathcal{P} = \begin{pmatrix} I_r\sigma_{k+1} & 0 \\ 0 & \mathcal{P}_2 \end{pmatrix}, \quad \mathcal{Q} = \begin{pmatrix} I_r\sigma_{k+1} & 0 \\ 0 & \mathcal{Q}_2 \end{pmatrix} \tag{3.20}$$

The *balancing transformation* (3.34) discussed in section 3.3 accomplishes this goal.

Clearly, *only* the pair $\mathcal{P}_2$, $\mathcal{Q}_2$ needs to be dilated. As in the suboptimal case, explicit formulae for $\hat{\mathbf{\Sigma}}$ can be obtained by using (3.6) of proposition 3.2. Partition A, B, C conformally with $\mathcal{P}$, $\mathcal{Q}$:

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

where $A_{11} \in \mathbb{R}^{r\times r}$, $B_1, C_1^* \in \mathbb{R}^r$ and I_r denotes the $r\times r$ identity matrix. The (1,1) block of the Lyapunov equations yields: $B_1B_1^* = C_1^*C_1$; this implies the existence of a unitary matrix U of size m, such that

$$B_1U = C_1^*, \quad UU^* = I_m$$

Using formulae (3.19) we construct an all-pass dilation of the subsystem (A_{22}, B_{21}, C_{12}); solving the corresponding equations (see (A.3)) we obtain

$$\begin{aligned} \hat{D} &= \sigma_{k+1}U \\ \hat{B} &= -\mathcal{Q}_2B_2 + C_2^*\hat{D} \\ \hat{C} &= (C_2\mathcal{P}_2 - \hat{D}B_2^*)(\Gamma_2^*)^{-1} \\ \hat{A} &= -A_{22}^* + C_2^*\hat{C} \end{aligned} \tag{3.21}$$

where $\Gamma_2 := \sigma_{k+1}^2 - \mathcal{P}_2\mathcal{Q}_2 \in \mathbb{R}^{(n-r)\times(n-r)}$. Hence unlike the sub-optimal case $\hat{D}$ is not arbirtary. In the single-input single-output case $\hat{D}$ is completely determined by the above relationship (is either $+1$ or -1) and hence there is a **unique** optimal approximant. This is not true for systems with more than one inputs and/or more than one outputs. Furthermore, in the one-step reduction case (i.e. $\sigma_{k+1} = \sigma_n$) the all-pass dilation system $\hat{\Sigma}$ has no antistable part, i.e. it is the optimal Hankel-norm approximant.

Corollary 3.2 **(a)** *In contrast to the sub-optimal case, in the optimal case, $\hat{\Sigma}$ has dimension r, and the dilated system Σ_e has dimension $2n-r$. Furthermore, the stable subsystem $\hat{\Sigma}_+$ has dimension $n-r$.* **(b)** *In case $\sigma_{k+1} = \sigma_n$, since $\hat{\Sigma} = \hat{\Sigma}_+$, $\Sigma - \hat{\Sigma}$ is stable and all-pass with norm σ_n.*

3.2.4 General case: parametrization of all sub-optimal approximants

Given the system Σ as in (3.1), let ϵ satisfy (2.25). The following rational matrix $\Theta(s)$ has size $(p+m)\times(p+m)$:

$$\begin{aligned}
\Theta(s) &:= I_{p+m} - C_\Theta(sI - A_\Theta)^{-1}\mathcal{Q}_\Theta^{-1}C_\Theta^* J \quad \text{where} \\
C_\Theta &:= \begin{pmatrix} C & 0 \\ 0 & B^* \end{pmatrix} \in \mathbb{R}^{(p+m)\times 2n} \\
A_\Theta &:= \begin{pmatrix} A & 0 \\ 0 & -A^* \end{pmatrix} \in \mathbb{R}^{2n\times 2n} \\
\mathcal{Q}_\Theta &:= \begin{pmatrix} \mathcal{Q} & \epsilon I_n \\ \epsilon I_n & \mathcal{P} \end{pmatrix} \in \mathbb{R}^{2n\times 2n}, \quad J := \begin{pmatrix} I_p & 0 \\ 0 & -I_m \end{pmatrix} \in \mathbb{R}^{(p+m)\times(p+m)}
\end{aligned} \tag{3.22}$$

Putting these expressions together we obtain:

$$\Theta(s) = \left(\begin{array}{c|c} \Theta_{11} & \Theta_{12} \\ \hline \Theta_{21} & \Theta_{22} \end{array}\right) := \left(\begin{array}{c|c} I_p + C(sI-A)^{-1}\Gamma^{-1}\mathcal{P}C^* & \epsilon C(sI-A)^{-1}\Gamma^{-1}B \\ \hline -\epsilon B^*(sI+A^*)^{-1}\Gamma^{-*}C^* & I_m - B^*(sI+A^*)^{-1}\mathcal{Q}\Gamma^{-1}B \end{array}\right) \tag{3.23}$$

By construction, Θ is J-unitary on the $j\omega$-axis, i.e. $\Theta^*(-j\omega)J\Theta(j\omega) = J$. Define

$$\begin{pmatrix} \Phi_1(\Delta) \\ \Phi_2(\Delta) \end{pmatrix} := \Theta(s) \begin{pmatrix} \Delta(s) \\ I_m \end{pmatrix} = \begin{pmatrix} \Theta_{11}(s)\Delta(s) + \Theta_{12}(s) \\ \Theta_{21}(s)\Delta(s) + \Theta_{22}(s) \end{pmatrix} \tag{3.24}$$

The proof of the following result can be found in chapter 24 of [8]. Recall main theorem 2.3.

Theorem 3.1 *$\hat{\Sigma}$ is an ϵ-all-pass dilation of Σ, if, and only if,*

$$H_{\Sigma} - H_{\hat{\Sigma}} = \Phi_1(\Delta)\Phi_2(\Delta)^{-1} \tag{3.25}$$

where $\Delta(s)$ is a $p\times m$ anti-stable contraction (all poles in the right-half plane and $\|\Delta(s)\|_\infty < 1$).

3.2.5 Error bounds of optimal and sub-optimal approximants

Given is the stable $m\times m$ system Σ, having Hankel singular values σ_i and multiplicity r_i, $i = 1, \cdots, q$ (see (2.21)). Let $\hat{\Sigma}$ be the ϵ-all-pass dilation of Σ, where $\epsilon = \sigma_q$ (the smallest singular value). Following proposition 3.2, the dimension of $\hat{\Sigma}_+$ is $n-q$, which implies that $\hat{\Sigma} = \hat{\Sigma}_+$, i.e. it is *stable*. Thus by 3.2(**b**), for a *one-step* reduction, the all-pass dilation system has *no* unstable poles; for simplicity we will use the notation

$$\mathbf{\Sigma}_q := \hat{\mathbf{\Sigma}}(\sigma_q)$$

Thus $\mathbf{\Sigma} - \mathbf{\Sigma}_q$ is *all-pass* of magnitude σ_q, and consequently:

$$\sigma_i(\mathbf{\Sigma}_q) = \sigma_i(\mathbf{\Sigma}),\ \ i = 1, \cdots, n - r_q \tag{3.26}$$

We now approximate $\mathbf{\Sigma}_q$ by $\mathbf{\Sigma}_{q-1}$ through a one-step optimal Hankel norm reduction. By successive application of the same procedure, we thus obtain a sequence of all-pass systems $\mathbf{\Sigma}_i$, $i = 1, \cdots, q$, and a system $\mathbf{\Sigma}_0$ consisting of the matrix D_0, such that the transfer function of $\mathbf{\Sigma}$ is decomposed as follows:

$$H(s) = D_0 + H_1(s) + \ \cdots\ + H_q(s) \tag{3.27}$$

where $H_k(s)$ is the transfer function of the stable, all-pass system $\mathbf{\Sigma}_k$, having dimension $2n - \sum_{i=k}^{q} r_i$, $k = 1, \cdots, q$; the dimension of the partial sums $\sum_{i=1}^{k} H_i(s)$, is equal to $\sum_{i=1}^{k} r_i$, $k = 1, 2, \cdots, q$. Thus (3.27) is the equivalent of the dyadic decomposition (2.4), for $\mathbf{\Sigma}$.

From the above decomposition we can derive the following upper bound for the $\mathcal{H}_\infty$ norm of $\mathbf{\Sigma}$. Assuming that $H_{\mathbf{\Sigma}}(\infty) = D = 0$:

$$\|H_{\mathbf{\Sigma}}(s) - D_0\|_\infty \leq \underbrace{\|H_1(s)\|_\infty}_{\sigma_1} + \cdots + \underbrace{\|H_k(s)\|_\infty}_{\sigma_k}$$

Evaluating the above expression at infinity yields:

$$\|D_0\|_2 \leq \sigma_1 + \ \cdots\ + \sigma_q$$

Thus, combining this inequality with $\|H_{\mathbf{\Sigma}}(s)\|_\infty \leq \|D_0\|_2 + \sigma_1 + \ \cdots\ + \sigma_q$, yields the bound given in lemma 3.1:

$$\|H_{\mathbf{\Sigma}}(s)\|_\infty \leq 2(\sigma_1 + \ \cdots\ + \sigma_q) \tag{3.28}$$

This bound can be sharpened by computing an appropriate $D_0 \in \mathbb{R}^{m\times m}$ as in (3.27):

$$\|H_{\mathbf{\Sigma}}(s) - D_0\|_\infty \leq \sigma_1 + \ \cdots\ + \sigma_q \tag{3.29}$$

Finally, we state a result of [14], on the Hankel singular values of the stable part $\mathbf{\Sigma}_{e+}$ of the all-pass dilation $\mathbf{\Sigma}_e$. It will be assumed for simplicity that each Hankel singular value has multiplicity one: $r_i = 1$. Assume that $\mathbf{\Sigma}_{e+} = \mathbf{\Sigma} - \hat{\mathbf{\Sigma}}_+$, where $\hat{\mathbf{\Sigma}}_+$ is an optimal Hankel approximant of $\mathbf{\Sigma}$ of dimension k; there holds: $\sigma_1(\mathbf{\Sigma}_{e+}) = \cdots = \sigma_{2k+1}(\mathbf{\Sigma}_{e+}) = \sigma_{k+1}(\mathbf{\Sigma})$, $\sigma_{2k+2}(\mathbf{\Sigma}_{e+}) = \sigma_1(\hat{\mathbf{\Sigma}}_-) \leq \sigma_{k+2}(\mathbf{\Sigma})$, $\cdots$, $\sigma_{n+k}(\mathbf{\Sigma}_{e+}) = \sigma_{n-k-1}(\hat{\mathbf{\Sigma}}_-) \leq \sigma_n(\mathbf{\Sigma})$. Using these inequalities we obtain an error bound for the $\mathcal{H}_\infty$-norm of the error of $\mathbf{\Sigma}$ with a degree k optimal approximant $\hat{\mathbf{\Sigma}}_+$. First, notice that

$$\delta := \|\mathbf{\Sigma}_- - D_0\|_\infty \leq \sigma_1(\hat{\mathbf{\Sigma}}_-) + \cdots + \sigma_{n-k-1}(\hat{\mathbf{\Sigma}}_-)$$

This implies $\|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}_+ - D_0\|_\infty \leq \sigma_{k+1}(\mathbf{\Sigma}) + \delta$. Combining all previous inequalities we obtain the bounds:

$$\boxed{\sigma_{k+1} \leq \|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}_+ - D_0\|_\infty \leq \sigma_{k+1}(\mathbf{\Sigma}) + \sigma_1(\hat{\mathbf{\Sigma}}_-) + \cdots + \sigma_{n-k-1}(\hat{\mathbf{\Sigma}}_-) \leq \sigma_{k+1}(\mathbf{\Sigma}) + \cdots + \sigma_n(\mathbf{\Sigma})} \tag{3.30}$$

The left-hand side inequality in (2.25), and the above finally yield upper and lower bounds on the $\mathcal{H}_\infty$-norm of the error:

$$\boxed{\sigma_{k+1} \leq \|\mathbf{\Sigma} - \hat{\mathbf{\Sigma}}_+\|_\infty \leq 2(\sigma_{k+1} + \cdots + \sigma_n)} \tag{3.31}$$

The first upper bound in (3.30) above is the tightest; but both a D-term and the singular values of the anti-stable $\hat{\Sigma}_-$ part of the all-pass dilation are needed. The second upper bound in (3.30) is the second-tightest; it only requires knowledge of the optimal D-term. Finally the upper bound in (3.31) is the less tight among these three. It is however the most useful, since it can be determined a priori, i.e. before the computation of the approximants, once the singular values of the original system Σ are known.

Remark 3.2 (a) *In a similar way a bound for the infinity norm of suboptimal approximants can be obtained. Given* Σ*, let* $\hat{\Sigma}$ *be an all-pass dilation satisfying (2.25). Then, similarly to (3.30), there holds:*

$$\sigma_{r+1} \leq \|\Sigma - \hat{\Sigma}_+\|_\infty \leq 2(\epsilon + \sigma_{r+1} + \cdots + \sigma_n)$$

(b) *The fact that for a one-step Hankel norm reduction* $\Sigma - \hat{\Sigma}$ *is all-pass, i.e.*

$$\|\Sigma - \hat{\Sigma}\|_H = \|\Sigma - \hat{\Sigma}\|_\infty = \sigma_q$$

shows that in this case $\hat{\Sigma}$ *is an optimal approximant not only in the Hankel norm but in the infinity norm as well. Thus for this special case, and despite propositions 2.1 and 3.1, the Hankel-norm approximation theory yields a solution for the optimal approximation problem in the 2-norm of the convolution operator* $\mathcal{S}_\Sigma$*; recall that this is the problem we initially wished to solve.* ■

3.3 Balanced realizations and balanced model reduction

A model reduction method which is closely related to the Hankel-norm approximation method is approximation by *balanced truncation.* This involves a particular realization of a linear system Σ given by (3.1), called *balanced realization* (the D matrix is irrelevant in this case). We start by explaining the rational behind this particular state-space realization.

The concept of balancing. Consider a stable system Σ, with positive definite reachability, observability grammians $\mathcal{P}$, $\mathcal{Q}$. It can be shown that the *minimal energy* required for the transfer of the state of the system from 0 to some final state x_r, is:

$$x_r^* \mathcal{P}^{-1} x_r \tag{3.32}$$

Similarly, the largest *observation energy* produced by observing the output, when the initial state of the system is x_o and the input is zero ($u = 0$), is equal to:

$$x_o^* \mathcal{Q} x_o \tag{3.33}$$

We conclude that the states which are difficult to reach, i.e. those which require a large amount of energy to reach, are in the span of the eigenvectors of the reachability grammian $\mathcal{P}$ corresponding to small eigenvalues. Moreover the states which are difficult to observe, i.e. those which yield small amounts of observation energy, are those which lie in the span of the eigenvectors of the observability grammian $\mathcal{Q}$ corresponding to small eigenvalues as well.

This observation suggests that one way to obtain reduced order models is by eliminating those states which are difficult to reach and/or difficult to observe. However, states which are difficult to reach may not be difficult to observe and vice-versa; as simple examples show, these properties are basis dependent. This suggests the need for a basis in which states which are difficult to reach are **simultaneously** difficult to observe, and vice-versa. From these considerations, the question arises: given a continuous- or discrete-time, stable system $\mathbf{\Sigma}$, does there exist a basis of the state space in which states which are *difficult to reach* are also *difficult to observe*?

The answer to this question is affirmative. The transformation which achieves this goal is called a *balancing transformation*. Under an equivalence transformation T (basis change in the state space: $\tilde{x} := Tx$, $\det T \neq 0$) we have: $\tilde{A} = TAT^{-1}$, $\tilde{B} = TB$, $\tilde{C} = CT^{-1}$, while the grammians are transformed as follows:

$$\tilde{\mathcal{P}} = T\mathcal{P}T^*, \;\; \tilde{\mathcal{Q}} = T^{-*}\mathcal{Q}T^{-1} \quad \Rightarrow \quad \tilde{\mathcal{P}}\tilde{\mathcal{Q}} = T\left(\mathcal{P}\mathcal{Q}\right)T^{-1}$$

The problem is to find T, $\det T \neq 0$, such that the transformed Grammians $\tilde{\mathcal{P}}$, $\tilde{\mathcal{Q}}$ are equal. This will ensure that the states which are difficult to reach are precisely those which are difficult to observe.

Definition 3.1 *The stable system* $\mathbf{\Sigma}$ *is* **balanced** *iff* $\mathcal{P} = \mathcal{Q}$. $\mathbf{\Sigma}$ *is* **principal-axis balanced** *iff*

$$\mathcal{P} = \mathcal{Q} = S := \mathrm{diag}\;(\sigma_1, \;\cdots\;, \sigma_n)$$

The existence of a balancing transformation is guaranteed.

Lemma 3.2 Balancing transformation. *Given the stable system* $\mathbf{\Sigma}$ *and the corresponding grammians* $\mathcal{P}$ *and* $\mathcal{Q}$*, let the matrices* R*,* U*, and* S *be defined by* $\mathcal{P} =: R^*R$ *and* $R\mathcal{Q}R^* =: US^2U^*$*. A (principal axis) balancing transformation is given by:*

$$T := S^{1/2}U^*R^{-*} \tag{3.34}$$

To verify that T is a balancing transformation, it follows by direct calculation that $T\mathcal{P}T^* = S$ and $T^{-*}\mathcal{Q}T^{-1} = S$. We also note that if the Hankel singular values are distinct (i.e. have multiplicity one) balancing transformations $\hat{T}$ are determined from T given above, up to multiplication by a *sign* matrix L, i.e. a diagonal matrix with ± 1 on the diagonal: $\hat{T} = LT$.

Model Reduction. Let $\mathbf{\Sigma}$ be balanced with grammians equal to S, and partition:

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_1 & 0 \\ 0 & S_2 \end{pmatrix}, \quad B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix}, \quad C = (C_1 \quad C_2) \tag{3.35}$$

The systems

$$\mathbf{\Sigma}_i := \left(\begin{array}{c|c} A_{ii} & B_i \\ \hline C_i & \end{array}\right), \;\; i = 1, 2$$

are called reduced order systems obtained form $\mathbf{\Sigma}$ by **balanced truncation**. These have certain guaranteed properties. However these properties are different for discrete- and continuous-time systems. Hence we state two theorems. For a proof see e.g. [25].

Theorem 3.2 Balanced truncation: continuous-time systems. *Given the stable (no poles in the closed right-half plane) continuous-time system* $\mathbf{\Sigma}$*, the reduced order systems* $\mathbf{\Sigma}_i$*,* $i = 1, 2$*, obtained by balanced truncation have the following properties:*

1. $\mathbf{\Sigma}_i$*,* $i = 1, 2$*, satisfy the Lyapunov equations* $A_{ii}S_i + S_iA_{ii}^* + B_iB_i^* = 0$*, and* $A_{ii}^*S_i + S_iA_{ii} + C_i^*C_i = 0$*. Furthermore* A_{ii}*,* $i = 1, 2$*, have no eigenvalues in the open right-half plane.*

2. *If S_1 and S_2 have no eigenvalues in common, both Σ_1 and Σ_2 have no poles on the imaginary axis, and are in addition reachable, observable, and balanced.*

3. *Let the distinct singular values of Σ be σ_i, with multiplicities m_i, $i = 1, \cdots, q$. Let Σ_1 have singular values σ_i, $i = 1, \cdots, k$, with the corresponding multiplicities m_i, $i = 1, \cdots, k$, $k < q$. The $\mathcal{H}_\infty$ norm of the difference between the full order system Σ and the reduced order system Σ_1 is upper-bounded by twice the sum of the neglected Hankel singular values:*

$$\|\Sigma - \Sigma_1\|_\infty \leq 2\,(\sigma_{k+1} + \;\cdots\; + \sigma_q) \tag{3.36}$$

If the smallest singular value is truncated, i.e. $S_2 = \sigma_q I_{r_q}$, equality holds.

Theorem 3.3 Balanced truncation: discrete-time systems. *Given the stable, discrete-time system (no poles in the complement of the open unit disc) Σ, the reduced order systems Σ_i obtained by balanced truncation have the following properties:*

1. *Σ_i, $i = 1, 2$ have no poles in the closed unit disc; these systems are in general not balanced.*

2. *If $\lambda_{min}(S_1) > \lambda_{max}(S_2)$, Σ_1 is in addition reachable and observable.*

3. *The h_∞ norm of the difference between full and reduced order models is upper-bounded by twice the sum of the neglected Hankel singular values, muliplicities included:*

$$\|\Sigma - \Sigma_1\|_\infty \leq 2\,\mathrm{trace}(S_2)$$

Remark 3.3 **(a)** *The last part of the above theorems says that if the neglected singular values are small, then the Bode plots of Σ and Σ_1 are guaranteed to be close in the $\mathcal{H}_\infty$ norm. The difference between part* **[3]** *for continuous- and discrete-time systems above, is that the muliplicities of the neglected singular values do not enter in the upper bound for continuous-time systems.*

(b) *Proposition 3.4 implies that the bilinear transformation between discrete- and continuous-time systems preserves balancing. 3.4 (see also the example 4.2 below).*

(c) *Let $\Sigma_{\rm hank}$, $\Sigma_{\rm bal}$ be the reduced order systems obtained by one step Hankel norm approximation, and one step balanced truncation, respectively. It can be shown that $\Sigma_{\rm hank} - \Sigma_{\rm bal}$ is all-pass with norm σ_q; it readily follows that*

$$\|\Sigma - \Sigma_{\rm bal}\|_\infty \leq 2\sigma_q \quad \text{and} \quad \|\Sigma - \Sigma_{\rm bal}\|_H \leq 2\sigma_q$$

A consequence of the above inequalities is that the error for reduction by balanced truncation can be upper-bounded, by means of the singular values of Σ as given in theorem 3.2. Furthermore, this bound is valid both for the $\mathcal{H}_\infty$ norm and for the Hankel norm.

(d) *Every linear, time-invariant, continuous-time and stable system Σ, can be expressed in* balanced canonical form. *In the generic case (distinct singular values) this canonical form is given in terms of $2n$ positive numbers, namely the singular values $\sigma_i > 0$, and some $b_i > 0$, as well as n signs $s_i = \pm 1$, $i = 1, \cdots, n$. The quantities $\lambda_i := s_i\sigma_i$, are called signed singular values of Σ; they satisfy: $H_{\Sigma}(0) = 2(\lambda_1 + \cdots + \lambda_n)$. For details on balanced canonical forms we refer to the work of Ober, e.g. [20].* ■

4 Examples

In this section we will illustrate the results presented above by means of 3 examples. The first deals with a simple 2nd order continuous-time system. The purpose is to compute the limit of suboptimal Hankel norm approximants as ϵ tends to one of the singular values. The second example discusses the approximation of a discrete-time system (3rd order FIR system) by balanced truncation and Hankel norm approximation. The section concludes with the approximation of the 4 classic analog filters (Butterworth, Chebyshev-1, Chebyshev-2, and Elliptic) by balanced truncation and Hankel norm approximation.

4.1 A simple continuous-time example

Consider the system $\mathbf{\Sigma}$ given by (3.1), where $n = 2$, $m = p = 1$ and

$$A = -\begin{pmatrix} \frac{1}{2\sigma_1} & \frac{1}{\sigma_1+\sigma_2} \\ \frac{1}{\sigma_1+\sigma_2} & \frac{1}{2\sigma_2} \end{pmatrix},\ B = \begin{pmatrix} 1 \\ 1 \end{pmatrix},\ C = \begin{pmatrix} 1 & 1 \end{pmatrix},\ D = 0$$

where $\sigma_1 > \sigma_2$. This system is in *balanced canonical form*; this means that the grammians are $\mathcal{P} = \mathcal{Q} = \text{diag}\,(\sigma_1, \sigma_2) =: S$; this canonical form is a special case of the forms discussed in [20].

We wish to compute the suboptimal Hankel norm approximants for $\sigma_1 > \epsilon > \sigma_2$. Then we will compute the limit of this family for $\epsilon \to \sigma_2$, and $\epsilon \to \sigma_1$, and show that the system obtained is indeed the optimal approximant. From (3.17) $\Gamma = \epsilon^2 - S^2 = \text{diag}\,(\epsilon^2 - \sigma_1^2, \epsilon^2 - \sigma_2^2)$; the inertia of Γ is $\{1, 0, 1\}$; furthermore from (3.19) :

$$\hat{A} = \begin{pmatrix} \frac{\epsilon-\sigma_1}{2\sigma_1(\epsilon+\sigma_1)} & \frac{\epsilon-\sigma_1}{(\sigma_1+\sigma_2)(\epsilon+\sigma_2)} \\ \frac{\epsilon-\sigma_2}{(\sigma_1+\sigma_2)(\epsilon+\sigma_1)} & \frac{\epsilon-\sigma_2}{2\sigma_2(\epsilon+\sigma_2)} \end{pmatrix},\ \hat{B} = \begin{pmatrix} \epsilon-\sigma_1 \\ \epsilon-\sigma_2 \end{pmatrix},\ \hat{C} = \begin{pmatrix} \frac{-1}{\epsilon+\sigma_1} & \frac{-1}{\epsilon+\sigma_2} \end{pmatrix},\ \hat{D} = \epsilon$$

Since the inertia of $\hat{A}$ is equal to the inertia of $-\Gamma$, $\hat{A}$ has one stable and one unstable poles (this can be checked directly by noticing that the determinant of $\hat{A}$ is negative). As $\epsilon \to \sigma_2$ we obtain

$$\hat{A} = \begin{pmatrix} \frac{\sigma_2-\sigma_1}{2\sigma_1(\sigma_1+\sigma_2)} & \frac{\sigma_2-\sigma_1}{2\sigma_2(\sigma_1+\sigma_2)} \\ 0 & 0 \end{pmatrix},\ \hat{B} = \begin{pmatrix} \sigma_2-\sigma_1 \\ 0 \end{pmatrix},\ \hat{C} = \begin{pmatrix} \frac{-1}{\sigma_1+\sigma_2} & \frac{-1}{2\sigma_2} \end{pmatrix},\ \hat{D} = \sigma_2$$

This system is not reachable but observable (i.e. there is a pole-zero cancellation in the transfer function). A state space representation of the reachable and observable subsystem is:

$$\bar{A} = \frac{\sigma_2-\sigma_1}{2\sigma_1(\sigma_1+\sigma_2)},\quad \bar{B} = \sigma_2-\sigma_1,\quad \bar{C} = \frac{-1}{\sigma_1+\sigma_2},\quad \bar{D} = \sigma_2$$

The formulae (3.19) depend on the choice of $\hat{D}$. If we choose it to be $-\epsilon$, the limit still exists and gives a realization of the optimal system which is equivalent to $\bar{A}, \bar{B}, \bar{C}, \bar{D}$ given above.

Finally, if $\epsilon \to \sigma_1$, after a pole-zero cancellation, we obtain the following reachable and observable approximant:

$$\bar{A} = \frac{\sigma_1-\sigma_2}{2\sigma_1(\sigma_1+\sigma_2)},\quad \bar{B} = \sigma_1-\sigma_2,\quad \bar{C} = \frac{-1}{\sigma_1+\sigma_2},\quad \bar{D} = \sigma_1$$

This is the best anti-stable approximant of Σ, i.e. the Nehari solution (see remark 2.3**(a)**). ■

4.2 A simple discrete-time example

In this section we will consider a 3rd order discrete-time FIR system described by the transfer function:

$$H(z)=\frac{z^2+1}{z^3} \tag{4.1}$$

We will first consider approximation by balanced truncation. The issue here is to examine balanced truncation first in discrete-time, and then in continuous-time, using the bilinear transformation of section 3.1.6, and compare the results. Subsequently, Hankel norm approximation will be investigated.

I. Approximation by balanced truncation. A balanced realization of this system is given by

$$\mathbf{\Sigma}_d := \left(\begin{array}{c|c} F & G \\ \hline H & J \end{array}\right) = \left(\begin{array}{ccc|c} 0 & \alpha & 0 & \beta \\ \alpha & 0 & -\alpha & 0 \\ 0 & \alpha & 0 & -\gamma \\ \hline -\beta & 0 & -\gamma & 0 \end{array}\right),\ \alpha = 5^{-\frac{1}{4}},\ \beta = \frac{\sqrt{3\sqrt{5}+5}}{\sqrt{10}},\ \gamma = \frac{\sqrt{3\sqrt{5}-5}}{\sqrt{10}}$$

where the reachability and observability grammians are equal and diagonal

$$\mathcal{P}=\mathcal{Q}=S=\mathrm{diag}\,(\sigma_1,\sigma_2,\sigma_3),\ \sigma_1=\frac{\sqrt{5}+1}{2},\ \sigma_2=1,\ \sigma_3=\frac{\sqrt{5}-1}{2}$$

The second and first order balanced truncated systems are

$$\mathbf{\Sigma}_{d,2} = \left(\begin{array}{c|c} F_2 & G_2 \\ \hline H_2 & J_2 \end{array}\right) = \left(\begin{array}{cc|c} 0 & \alpha & \beta \\ \alpha & 0 & 0 \\ \hline -\beta & 0 & 0 \end{array}\right),\quad \mathbf{\Sigma}_{d,1} = \left(\begin{array}{c|c} F_1 & G_1 \\ \hline H_1 & J_1 \end{array}\right) = \left(\begin{array}{c|c} 0 & \beta \\ \hline -\beta & 0 \end{array}\right)$$

Notice that $\mathbf{\Sigma}_{d,2}$ is *balanced*, but has singular values which are *different* from σ_1, σ_2. $\mathbf{\Sigma}_{d,1}$ is also balanced since $G_1=-H_2$ but its grammians are not equal to σ_1.

Let $\mathbf{\Sigma}_c$ denote the continuous time system obtained from $\mathbf{\Sigma}_d$ by means of the bilinear transformation described in section 3.1.6:

$$\mathbf{\Sigma}_c = \left(\begin{array}{c|c} A & B \\ \hline C & D \end{array}\right) = \left(\begin{array}{ccc|c} -1-2\alpha^2 & 2\alpha & 2\alpha^2 & \delta_+ \\ 2\alpha & -1 & -2\alpha & -\sqrt{2} \\ -2\alpha^2 & -2\alpha & -1+2\alpha^2 & \delta_- \\ \hline -\delta_+ & \sqrt{2} & \delta_- & 2 \end{array}\right)$$

where $\delta_\pm=\sqrt{2}(\alpha\pm\beta)$. Notice that $\mathbf{\Sigma}_c$ is balanced. We now compute first and second reduced order systems $\mathbf{\Sigma}_{c,1}$, $\mathbf{\Sigma}_{c,2}$ by truncating $\mathbf{\Sigma}_c$:

$$\mathbf{\Sigma}_{c,2} = \left(\begin{array}{c|c} A_2 & B_2 \\ \hline C_2 & D_2 \end{array}\right) = \left(\begin{array}{cc|c} -1-2\alpha^2 & 2\alpha & \delta_+ \\ 2\alpha & -1 & -\sqrt{2} \\ \hline -\delta_+ & \sqrt{2} & 2 \end{array}\right)$$

$$\mathbf{\Sigma}_{c,1} = \left(\begin{array}{c|c} A_1 & B_1 \\ \hline C_1 & D_1 \end{array}\right) = \left(\begin{array}{c|c} -1-2\alpha^2 & \delta_+ \\ \hline -\delta_+ & 2 \end{array}\right)$$

Let $\bar{\mathbf{\Sigma}}_{d,2}$, $\bar{\mathbf{\Sigma}}_{d,1}$ be the discrete-time systems obtaining by transforming $\mathbf{\Sigma}_{c,2}$, $\mathbf{\Sigma}_{c,1}$ back to discrete-time:

$$\bar{\mathbf{\Sigma}}_{d,2} = \left(\begin{array}{c|c} \bar{F}_2 & \bar{G}_2 \\ \hline \bar{H}_2 & \bar{J}_2 \end{array}\right) = \left(\begin{array}{cc|c} 0 & \alpha & \beta \\ \alpha & \alpha^2 & -\alpha\gamma \\ \hline -\beta & \alpha\gamma & -\gamma^2 \end{array}\right)$$

$$\bar{\mathbf{\Sigma}}_{d,1} = \left(\begin{array}{c|c} \bar{F}_1 & \bar{G}_1 \\ \hline \bar{H}_1 & \bar{J}_1 \end{array}\right) = \left(\begin{array}{c|c} -\sigma_3/2 & (\alpha+\beta)\sigma_3/2\alpha^2 \\ \hline -(\alpha+\beta)\sigma_3/2\alpha^2 & 2-(\alpha^2+\beta)^2/(1+\alpha^2) \end{array}\right)$$

The conclusion is that $\bar{\mathbf{\Sigma}}_{d,2}$, $\bar{\mathbf{\Sigma}}_{d,1}$ are balanced and different from $\mathbf{\Sigma}_{d,2}$, $\mathbf{\Sigma}_{d,1}$. It is interesting to notice that the singular value of $\mathbf{\Sigma}_{d,1}$ is $\beta^2 = \frac{1+\sigma_1}{\sqrt{5}}$, while that of $\bar{\mathbf{\Sigma}}_{d,1}$ is σ_1; β^2 satisfies $\sigma_2 < \beta^2 < \sigma_1$. Furthermore the singular values of $\mathbf{\Sigma}_{d,2}$ are $5\beta^2/4$, $\sqrt{5}\beta^2/4$ which satisfy the following interlacing inequalities:

$$\sigma_3 < \sqrt{5}\beta^2/4 < \sigma_2 < 5\beta^2/4 < \sigma_1$$

The numerical values of the quantities above are: $\sigma_1 = 1.618034$, $\sigma_3 = 0.618034$, $\alpha = 0.66874$, $\beta = 1.08204$, $\gamma = 0.413304$, $\delta_+ = 2.47598$, $\delta_- = .361241$. ■

II. Hankel-norm approximation. The Hankel operator of the system described by (4.1) is:

$$\mathcal{H} = \begin{pmatrix} 1 & 0 & 1 & 0 & \\ 0 & 1 & 0 & 0 & \cdots \\ 1 & 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 & \\ & \vdots & & & \ddots \end{pmatrix}$$

The SVD of the 3×3 principal submatrix of $\mathcal{H}$ is:

$$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\sigma_1}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_3}{\sqrt{5}}} \\ 0 & 1 & 0 \\ \sqrt{\frac{\sigma_3}{\sqrt{5}}} & 0 & -\sqrt{\frac{\sigma_1}{\sqrt{5}}} \end{pmatrix} \begin{pmatrix} \sigma_1 & 0 & 0 \\ 0 & \sigma_2 & 0 \\ 0 & 0 & \sigma_3 \end{pmatrix} \begin{pmatrix} \sqrt{\frac{\sigma_1}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_3}{\sqrt{5}}} \\ 0 & 1 & 0 \\ -\sqrt{\frac{\sigma_3}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_1}{\sqrt{5}}} \end{pmatrix} \tag{4.2}$$

where σ_i, $i = 1, 2, 3$, are as given earlier. It is tempting to conjecture that the optimal second order approximant is obtained by setting $\sigma_3 = 0$ in (4.2). The problem with this procedure is that the resulting approximant does *not* have Hankel structure.

To compute the optimal approximant, the system is first transformed to a continuous-time system using the transformation of subsection 3.1.6; we obtain the transfer function:

$$H_c(s) = H_d\left(\frac{1+s}{1-s}\right) = \frac{2(s^3 - s^2 + s - 1)}{(s+1)^3}$$

where H_d is the transfer function (4.1). Applying the theory discussed in subsection 3.2.3, we obtain the following second order continuous-time optimal approximant:

$$H_{c,2}(s) = \frac{-(s^2-1)}{(1-\sigma_3)s^2 + 2\sigma_1 s + (1-\sigma_3)}$$

Again using the transformation of subsection 3.1.6 we obtain the following discrete-time optimal approximant:

$$H_{d,2}(z) = H_{c,2}\left(\frac{z-1}{z+1}\right) = \frac{z}{z^2 - \sigma_3}$$

Notice that the optimal approximant is *not* an FIR system. It has poles at $\pm\sqrt{\sigma_3}$. Furthermore, the error

$$H_d(z) - H_{d,2}(z) = \sigma_3 \left[\frac{1-\sigma_3 z^2}{z^3(z^2-\sigma_3)} \right]$$

is all pass with magnitude equal to σ_3, on the unit circle (as predicted by corollary 3.2(**b**)). The corresponding optimal Hankel matrix of rank 2 is:

$$\hat{\mathcal{H}} = \begin{pmatrix} 1 & 0 & \sigma_3 & 0 & \sigma_3^2 & \\ 0 & \sigma_3 & 0 & \sigma_3^2 & 0 & \cdots \\ \sigma_3 & 0 & \sigma_3^2 & 0 & \sigma_3^3 & \\ 0 & \sigma_3^2 & 0 & \sigma_3^3 & 0 & \\ \sigma_3^2 & 0 & \sigma_3^3 & 0 & \sigma_3^4 & \\ & \vdots & & & & \ddots \end{pmatrix}$$

In this particular case the 3×3 principal submatrix of $\hat{\mathcal{H}}$ is also an optimal approximant of the corresponding submatrix of $\mathcal{H}$.

$$\begin{pmatrix} 1 & 0 & \sigma_3 \\ 0 & \sigma_3 & 0 \\ \sigma_3 & 0 & \sigma_3^2 \end{pmatrix} = \begin{pmatrix} \sqrt{\frac{\sigma_1}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_3}{\sqrt{5}}} \\ 0 & 1 & 0 \\ \sqrt{\frac{\sigma_3}{\sqrt{5}}} & 0 & -\sqrt{\frac{\sigma_1}{\sqrt{5}}} \end{pmatrix} \begin{pmatrix} 1+\sigma_3^2 & 0 & 0 \\ 0 & \sigma_3 & 0 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{\frac{\sigma_1}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_3}{\sqrt{5}}} \\ 0 & 1 & 0 \\ -\sqrt{\frac{\sigma_3}{\sqrt{5}}} & 0 & \sqrt{\frac{\sigma_1}{\sqrt{5}}} \end{pmatrix} \tag{4.3}$$

Notice that the above decomposition can be obtained from (4.2) by making use of the freedom mentioned in (2.7). Finally it is readily checked that the Hankel matrix consisting of 1 as $(1,1)$ entry and 0 everywhere else, is the optimal approximant of $\mathcal{H}$ of rank one. The dyadic decomposition (3.27) of H is:

$$H(z) = H_1(z) + H_2(z) + H_3(z) = \sigma_1 \left[\frac{1}{z}\right] + \sigma_2 \left[\frac{1-\sigma_3 z^2}{z(z^2-\sigma_3)} \right] + \sigma_3 \left[\frac{-1+\sigma_3 z^2}{z^3(z^2-\sigma_3)} \right]$$

Notice that each H_i is σ_i-all-pass, and the degree of H_1 is one, that of $H_1 + H_2$ is two and finally that of all three summands is three. ■

4.3 A higher order example

In our last example, we will approximate four well-known types of analog filters by means of balanced truncation and Hankel norm approximation. These are:

1. $\Sigma_{\rm B}$ - Butterworth:
2. $\Sigma_{\rm C1}$ - Chebyshev-1: 1% ripple in the pass band (PB)
3. $\Sigma_{\rm C2}$ - Chebyshev-2: 1% ripple in the stop band (SB)
4. $\Sigma_{\rm E}$ - Elliptic: 0.1% ripple both in the PB and in the SB

In each case we will consider 20-th order low pass filters, with pass band gain equal to 1, and cut-off frequency normalized to 1. Figure 2 shows the Hankel singular values of the full-order models. It follows from these plots that in order to obtain roughly comparable approximation errors, $\Sigma_{\rm B}$ and $\Sigma_{\rm C2}$ will have to be approximated by systems of lower order than $\Sigma_{\rm C1}$ and $\Sigma_{\rm E}$; we thus choose to approximate $\Sigma_{\rm B}$, $\Sigma_{\rm C2}$ by *8th order* models, and $\Sigma_{\rm C1}$, $\Sigma_{\rm E}$ by *10th order* models. It is interesting to observe that for the Chebyshev-2 filter the difference of the singular values: $\sigma_{13} - \sigma_{20}$, is of the order 10^{-7}. Thus $\mathbf{\Sigma}_{\mathrm{C2}}$ has an (approximate) 8^{th}-order all-pass subsystem of magnitude .05. Consequently, $\mathbf{\Sigma}_{\mathrm{C2}}$ cannot be approximated with systems of order 13 through 19. Similarly, the Chebyshev-1 filter has an all-pass subsystem of order 3; consequently approximations of order 1,2,3 are not possible.

In the sequel, the subscript "bal" stands for approximation by balanced truncation; the subscript "hank" stands for optimal Hankel norm approximation; "FOM" stands for "full order model"; "ROM" stands for "reduced order model".

Figure 3 gives the amplitude Bode plots of the error systems and tabulates their $\mathcal{H}_\infty$ norms and upper bounds. We observe that the $10th$ order Hankel-norm approximants of $\mathbf{\Sigma}_{\mathrm{C1}}$ and $\mathbf{\Sigma}_{\mathrm{E}}$ are not very good in the SB; one way for improving them is to increase the approximation order; another is to compute weighted approximants (see e.g. [24]). In the table below more detailed bounds for the approximants will be given in the case where the approximant contains an optimal D-term. "Bound 1" is the first expression on the right-hand side of (3.30), and "Bound 2" is the second expression on the right-hand side of the same expression:

$$\text{Bound 11} := \sigma_9(\mathbf{\Sigma}) + \sigma_1(\hat{\mathbf{\Sigma}}_-) + \cdots + \sigma_{11}(\hat{\mathbf{\Sigma}}_-) \leq \sigma_9(\mathbf{\Sigma}) + \cdots + \sigma_{20}(\mathbf{\Sigma}) =: \text{Bound 12}$$

$$\text{Bound 21} := \sigma_{11}(\mathbf{\Sigma}) + \sigma_1(\hat{\mathbf{\Sigma}}_-) + \cdots + \sigma_9(\hat{\mathbf{\Sigma}}_-) \leq \sigma_{11}(\mathbf{\Sigma}) + \cdots + \sigma_{20}(\mathbf{\Sigma}) =: \text{Bound 22}$$

	Optimal D	σ_9	$\Vert\mathbf{\Sigma} - \Sigma_{\mathrm{hank}}\Vert_\infty$	Bound 11	Bound 12
BUTTERWORTH	0.0384	0.0384	0.0389	0.0390	0.0517
CHEBYSHEV-2	0.1010	0.0506	0.1008	0.5987	0.6008
	Optimal D	σ_{11}	$\Vert\mathbf{\Sigma} - \Sigma_{\mathrm{hank}}\Vert_\infty$	Bound 21	Bound 22
CHEBYSHEV-1	0.2988	0.3374	0.4113	0.5030	0.9839
ELLIPTIC	0.2441	0.2458	0.2700	0.3492	0.7909

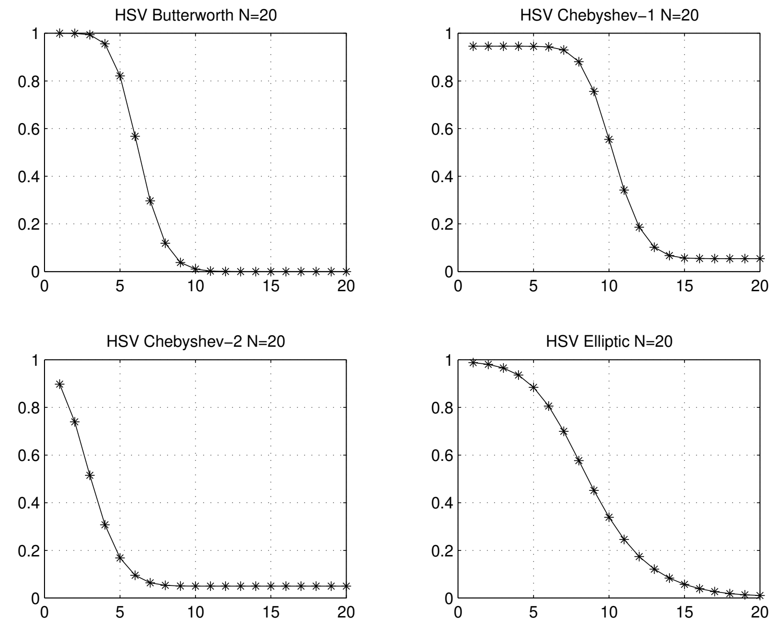
5 Epilogue

The computational algorithms that emerged from balancing and Hankel norm model reduction, have found their way to software packages. The matlab toolboxes (i) Robust control toolbox [9] and (ii) μ-toolbox [7], contain m-files which address the approximation problems discussed above. Two such m-files are **sysbal** and **hankmr**; the first is used for balancing and balanced truncation, while the second is used for optimal Hankel norm approximation (including the computation of the anti-stable part of the all-pass dilation).

We will conclude with a brief discussion of some papers which were written since Glover's seminal paper [14], namely: [17], [13], [11], [16], [24], [6], [18], [3].

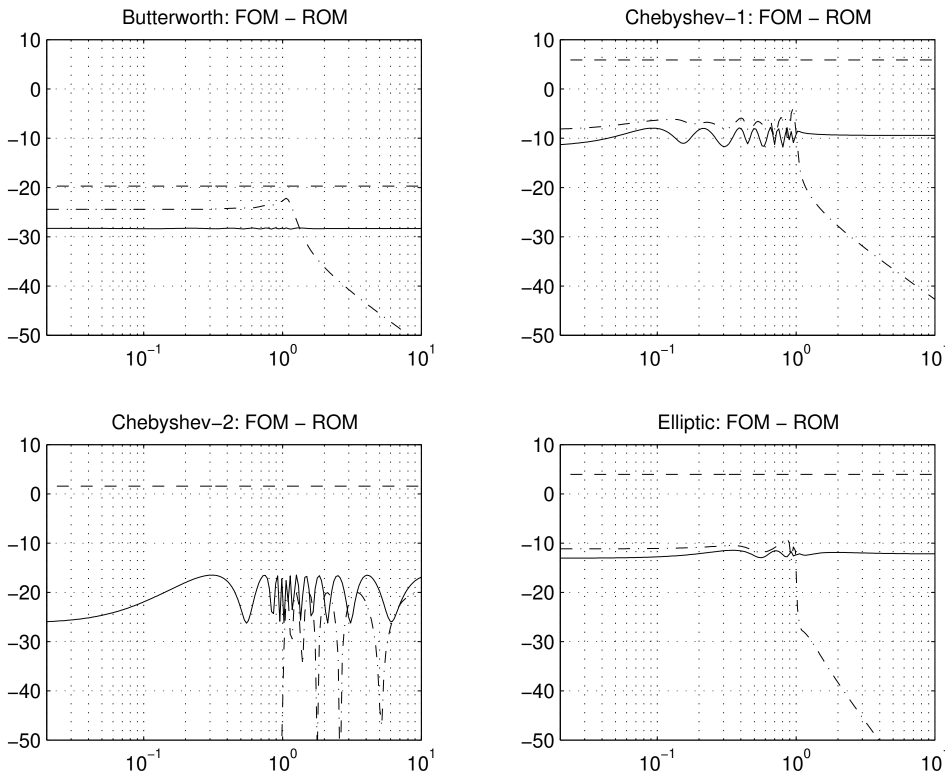
Hankel norm approximation or balanced truncation can be applied to stable systems. The approximants are guaranteed to be close to the original system, within the bounds given in section 3.2.5; the important aspect of this theory is that these bounds can be computed a priori, i.e. before computing the approximant. The bounds are in terms of the $\mathcal{H}_\infty$ norm which is a well-defined measure of distance between stable systems and has a direct relevance in feedback control theory.

When the system is unstable however, Hankel norm approximation, may not be the right thing to do, since it would not predict closeness in any meaningful way. To bypass this difficulty,



	Butterworth	Chebyshev-1	Chebyshev-2	Elliptic
σ_1	9.9998e-01	9.5000e-01	8.9759e-01	9.8803e-01
σ_2	9.9952e-01	9.5000e-01	7.3975e-01	9.8105e-01
σ_3	9.9376e-01	9.5000e-01	5.1523e-01	9.6566e-01
σ_4	9.5558e-01	9.4996e-01	3.0731e-01	9.3602e-01
σ_5	8.2114e-01	9.4962e-01	1.6837e-01	8.8469e-01
σ_6	5.6772e-01	9.4710e-01	9.5205e-02	8.0582e-01
σ_7	2.9670e-01	9.3362e-01	6.3885e-02	6.9997e-01
σ_8	1.1874e-01	8.8268e-01	5.3342e-02	5.7677e-01
σ_9	3.8393e-02	7.5538e-01	5.0643e-02	4.5166e-01
σ_{10}	1.0402e-02	5.5133e-01	5.0103e-02	3.3873e-01
σ_{11}	2.3907e-03	3.3740e-01	5.0014e-02	2.4576e-01
σ_{12}	4.6611e-04	1.8209e-01	5.0002e-02	1.7413e-01
σ_{13}	7.6629e-05	9.8132e-02	5.0000e-02	1.2135e-01
σ_{14}	1.0507e-05	6.3256e-02	5.0000e-02	8.3613e-02
σ_{15}	1.1813e-06	5.2605e-02	5.0000e-02	5.7195e-02
σ_{16}	1.0622e-07	5.0362e-02	5.0000e-02	3.9014e-02
σ_{17}	7.3520e-09	5.0036e-02	5.0000e-02	2.6729e-02
σ_{18}	3.6804e-10	5.0003e-02	5.0000e-02	1.8650e-02
σ_{19}	1.1868e-11	5.0000e-02	5.0000e-02	1.3613e-02
σ_{20}	1.8521e-13	5.0000e-02	5.0000e-02	1.0869e-02

Figure 2: Analog Filter Approximation: Hankel singular values (curves and numerical values)



	$\mathcal{H}_\infty$ Norm of the Error and Bounds			
	σ_9	$\lVert\mathbf{\Sigma}-\mathbf{\Sigma}_{\mathrm{hank}}\rVert_\infty$	$\lVert\mathbf{\Sigma}-\mathbf{\Sigma}_{\mathrm{bal}}\rVert_\infty$	$2(\sigma_9+\cdots+\sigma_{20})$
Butterworth	0.0383	0.0388	0.0779	0.1035
Chebyshev-2	0.0506	0.1008	0.0999	1.2015
	σ_{11}	$\lVert\mathbf{\Sigma}-\mathbf{\Sigma}_{\mathrm{hank}}\rVert_\infty$	$\lVert\mathbf{\Sigma}-\mathbf{\Sigma}_{\mathrm{bal}}\rVert_\infty$	$2(\sigma_{11}+\cdots+\sigma_{20})$
Chebyshev-1	0.3374	0.4113	0.6508	1.9678
Elliptic	0.2457	0.2700	0.3595	1.5818

Figure 3: Analog Filter Approximation: Bode plots of the error systems for model reduction by optimal Hankel-norm approximation (continuous curves), balanced truncation (dash-dot curves), and the upper bound (3.31) (dash-dash curves). The plots are followed by a table comparing the peak values of these Bode plots with the lower and upper bounds predicted by the theory.

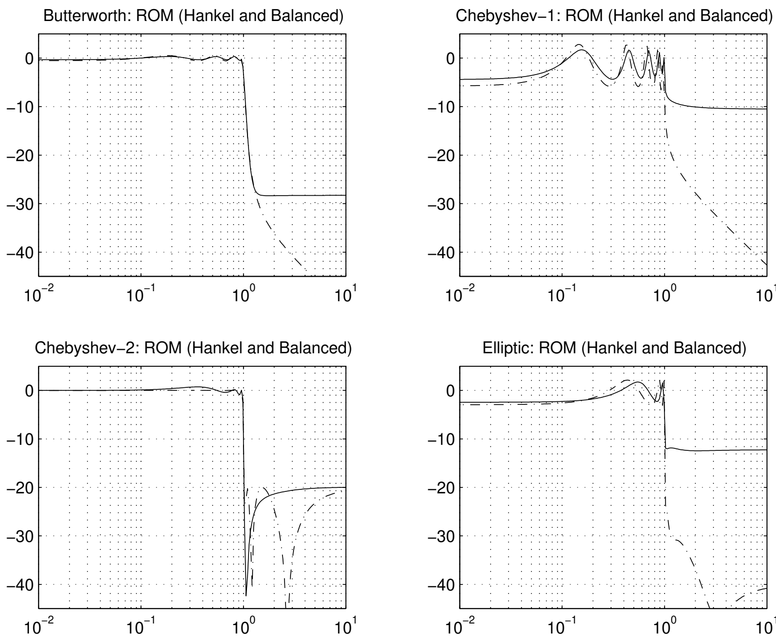


Figure 4: Analog Filter Approximation: Bode plots of the reduced order models obtained by balanced truncation (dash-dot curves) and Hankel norm approximation (continuous curves)

[17] proposes approximating unstable systems (in the balanced and Hankel norm sense) using the (normalized) *coprime factors*, derived from the transfer function Another way to go about reducing unstable systems is by adopting the *gap metric* for measuring the distance between two systems; the gap is a natural measure of distance in the context of feedback control. For details on this approach we refer to the work by Georgiou and Smith [13], and references therein.

The second paper [11] authored by Fuhrmann, uses the author's *polynomial models* for a detailed analysis of the Hankel operator with given (scalar) rational symbol (i.e. single-input single-output systems). Fuhrmann's computations follow a different approach than in [14]. The Schmidt pairs (singular vectors) are explicitly computed and hence the optimal Hankel approximants are obtained. This analysis yields new insights into the problem; we refer to [11], theorem 8.1 on page 184, which shows that the decomposition of the transfer function in the basis provided by the Schmidt pairs of the Hankel operator provides the balanced realization of the associated state space representation. Another advantage is that it suggests the correct treatment of the polynomial equation (3.12), which yields a set of n linear equations, instead of $2n-1$ as in (3.13). Further developments along the same lines are given in [12].

Weighted Hankel norm approximation was introduced in [16]. Reference [24] presents a new frequency weighted balanced truncation method. It comes with explicit $\mathcal{L}_\infty$ norm error bounds. Furthermore the weighted Hankel norm problem with anti-stable weighting is solved. These results are applied to $\mathcal{L}_\infty$ norm model reduction by means of Hankel norm approximation.

Paper [6] discusses a rational interpolation approach to Hankel norm approximation.

The next paper [18] addresses the issue of balanced truncation for second-order form linear systems. These are systems which arise in applications and which are modeled by position and velocity vectors. Thus in simplifying, if one decides to eliminate the variable q_i, the derivative $\dot{q}_i$ should be eliminated as well, and vice versa. Towards this goal new sets of invariants are introduced and new grammians as well. However, no bounds on the $\mathcal{H}_\infty$ norm of the error are provided.

The question whether the bounds (3.31) and (3.36) are tight arises. It can be shown that this property holds for systems having zeros which interlace the poles; an example of such systems are RC circuits (this result follows from (3.31) and remark 3.3**(d)**, after noticing that the Hankel operator in this case is positive definite). For an elementary exposition of this result we refer to [22]; for a more abstract account, see [19].

The problems of approximating linear operators in the 2-induced norm, which are (1) finite dimensional, unstructured, and (2) infinite dimensional structured (Hankel), have been solved. The solutions of these two problems exhibit striking similarities. These similarities suggest the search of a *unifying framework* for the approximation of linear operators in the 2-induced norm. For details see [4]. In particular the problem of approximating a finite dimensional Hankel matrix remains largely unknown. Some partial results are provided in [3] which relate to results in [11].

References

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

[2] V.M. Adamjan, D.Z. Arov, and M.G. Krein, *Infinite block Hankel matrices and related extension problems*, American Math. Society Transactions, **111**: 133-156 (1978).

[3] A.C. Antoulas, *On the approximation of Hankel matrices*, in Operators, Systems and Linear Algebra, edited by U. Helmke, D. Prätzel-Wolters and E. Zerz, pages 17-23, Teubner Verlag, Stuttgart (1997).

[4] A.C. Antoulas, *Approximation of linear operators in the 2-norm*, Linear Algebra and its Applications, Special Issue on *Challenges in Matrix Theory*, (1998).

[5] A.C. Antoulas, *Lectures on optimal approximation of linear systems*, Draft (1998).

[6] T. Auba and Y. Funabashi, *Interpolation approach to Hankel norm model reduction for rational multi-input multi-output systems*, IEEE Transactions on Circuits and Systems-I: Fundamental Theory and Applications, **43**: 987-995 (1996).

[7] G.J. Balas, J.C. Doyle, K. Glover, A. Packard, and R. Smith, *μ-Analysis and Synthesis Toolbox*, The MathWorks (1993).

[8] J.A. Ball, I. Gohberg, and L. Rodman, *Interpolation of rational matrix functions*, Birkhäuser Verlag (1990).

[9] R.Y. Chiang and M.G. Safonov, *Robust control toolbox*, The MathWorks Inc., (1992).

[10] P.A. Fuhrmann, *Linear systems and operators in Hilbert space*, McGraw Hill (1981).

[11] P.A. Fuhrmann, *A polynomial approach to Hankel norm and balanced approximations*, Linear Algebra and its Applications, **146**: 133-220 (1991).

[12] P.A. Fuhrmann and R. Ober, *A functional approach to LQG balancing*, International Journal of Control, **57**: 627–741 (1993).

[13] T.T. Georgiou and M.C. Smith, *Approximation in the gap metric: upper and lower bounds*, IEEE Trans. on Automatic Control, **38**: 946-951 (1993).

[14] K. Glover, *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds*, International Journal of Control, **39**: 1115-1193 (1984).

[15] M. Green and D.J.N. Limebeer, *Linear robust control*, Prentice Hall, Upper Saddle River, NJ, (1995).

[16] G.A. Latham and B.D.O. Anderson, *Frequency-weighted optimal Hankel norm approximation of stable transfer functions*, Systems and Control Letters, **5**: 229-236 (1985).

[17] D.G. Meyer, *A fractional approach to model reduction*, Proc. of American Control Conference, Atlanta, pages 1041-1047 (1988).

[18] D.G. Meyer and S. Srinivasan, *Balancing and model reduction for second-order form linear systems*, IEEE Trans. on Automatic Control, **41**: 1632-1644 (1996).

[19] R. Ober, *On Stieltjes functions and Hankel operators*, Systems and Control Letters, **27**: 275-278 (1996).

[20] R. Ober, *Balanced parametrization of classes of linear systems*, SIAM J. Control and Optimization, **29**: 1251-1287 (1991).

[21] W.J. Rugh, *Linear system theory*, Second Edition, Prentice Hall, Upper Saddle River, NJ, (1996).

[22] B. Srinivasan and P. Myszkorowski, *Model reduction of systems zeros interlacing the poles*, Systems and Control Letters, **30**: 19-24 (1997).

[23] G.W. Stewart and J. Sun, *Matrix perturbation theory*, Academic Press (1990).

[24] K. Zhou, *Frequency-weighted $\mathcal{L}_\infty$ norm and optimal Hankel norm model reduction*, IEEE Trans. on Automatic Control, **40**: 1687-1699 (1995).

[25] K. Zhou, J.C. Doyle, and K. Glover, *Robust and optimal control*, Prentice Hall, Upper Saddle River, NJ, (1996).

A Appendix

The equations (3.6) needed to derive (3.19):

$$\begin{pmatrix} A^* & 0 \\ 0 & \hat{A}^* \end{pmatrix}\begin{pmatrix} \mathcal{Q} & I_n \\ I_n & -\Gamma^{-1}\mathcal{P} \end{pmatrix}+\begin{pmatrix} \mathcal{Q} & I_n \\ I_n & -\Gamma^{-1}\mathcal{P} \end{pmatrix}\begin{pmatrix} A & 0 \\ 0 & \hat{A} \end{pmatrix}+\begin{pmatrix} C^*C & -C^*\hat{C} \\ -\hat{C}^*C & \hat{C}^*\hat{C} \end{pmatrix}=\begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \tag{A.1}$$

$$\begin{pmatrix} \mathcal{Q} & I_n \\ I_n & -\Gamma^{-1}\mathcal{P} \end{pmatrix}\begin{pmatrix} B \\ \hat{B} \end{pmatrix}+\begin{pmatrix} C^* \\ -\hat{C}^* \end{pmatrix}(-\hat{D})=\begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{A.2}$$

and the equations needed to derive (3.21):

$$\begin{pmatrix} A_{11}^* & A_{21}^* & 0 \\ A_{12}^* & A_{22}^* & 0 \\ 0 & 0 & \hat{A}^* \end{pmatrix}\begin{pmatrix} \sigma I & 0 & 0 \\ 0 & \mathcal{Q}_2 & I \\ 0 & I & -\Gamma_2^{-1}\mathcal{P}_2 \end{pmatrix} + \begin{pmatrix} \sigma I & 0 & 0 \\ 0 & \mathcal{Q}_2 & I \\ 0 & I & -\Gamma_2^{-1}\mathcal{P}_2 \end{pmatrix}\begin{pmatrix} A_{11} & A_{12} & 0 \\ A_{21} & A_{22} & 0 \\ 0 & 0 & \hat{A} \end{pmatrix} + \tag{A.3}$$

$$+ \begin{pmatrix} C_1^*C_1 & C_1^*C_2 & -C_1^*\hat{C} \\ C_2^*C_1 & C_2^*C_2 & -C_2^*\hat{C} \\ -\hat{C}^*C_1 & -\hat{C}^*C_2 & \hat{C}^*\hat{C} \end{pmatrix}=0 \tag{A.4}$$

$$\begin{pmatrix} \sigma I & 0 & 0 \\ 0 & \mathcal{Q}_2 & I \\ 0 & I & -\Gamma_2^{-1}\mathcal{P}_2 \end{pmatrix}\begin{pmatrix} C_1^* \\ C_2^* \\ -\hat{C} \end{pmatrix}\left(-\hat{D}\right)=0 \tag{A.5}$$