Stony Brook University

Identification and Model Reduction of MIMO systems in Triangular Input Balanced Form

A Dissertation Presented

by

Xiao Yu

to

The Graduate School

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

December 2014

Stony Brook University

The Graduate School

Xiao Yu

We, the dissertation committee for the above candidate for the Doctor of Philosophy degree, hereby recommend acceptance of this dissertation

Andrew.P.Mullhaupt - Dissertation Advisor

Research Professor

Dept. of Applied Mathematics and Statistics, Stony Brook Univ.

Svetlozar T.Rachev - Chair Person of Defense

Frey Family Foundation Chair of Quantitative Finance

Dept. of Applied Mathematics and Statistics, Stony Brook Univ.

John D.Pinezich - Committee Member

Adjunct Professor

Dept. of Applied Mathematics and Statistics, Stony Brook Univ.

Kurt S.Riedel - External Committee Member

Portfolio Manager

Millennium Partners

This dissertation is accepted by the Graduate School

Charles Taber

Dean of the Graduate School

Abstract of the Dissertation

Identification and Model Reduction of MIMO systems in Triangular Input Balanced Form

by

Xiao Yu

Doctor of Philosophy

in

Applied Mathematics and Statistics

Stony Brook University

December 2014

Consider a discrete-time linear time invariant (LTI) d-dimensional innovations model,

$$z(t+1) = Az(t) + Bx(t), \quad (1)$$

$$y(t) = Cz(t) + x(t), \quad (2)$$

where $y(t)$ is a sequence of d-dimensional measurement vectors and $z(t)$ is a state vector of dimension n. The triangular input balanced (TIB) representation of the LTI system was introduced by A.Mullhaupt and K.Riedel [1] in 1995. In the Single-Input Single-Output (SISO) case, the TIB pair is uniquely determined by the poles of the system. However, the poles alone are not enough to fully characterize the Multi-Input Multi-Output (MIMO) TIB pair. We parametrize MIMO transfer functions in terms of data of the Schur tangential algorithm. The inner part appears as a Blaschke-Potapov factorization. We relate these parameters to TIB and lattice realizations, and use this correspondence to construct novel methods for model reduction and identification.

To my parents

Contents

List of Tables

List of Figures

Acknowledgement

It has been four and half years since I came to the States from China to pursue my doctoral degree, I am grateful that I have received tremendous help from numerous people.

Foremost on this list is my advisor, Professor Andrew Mullhaupt, who gave me a good problem to solve, provided guidance on my research always patience. His insights showed me what an interesting field the matrix analysis can be and how it is related to almost every field of modern applied mathematics. He is a great educator, he taught me not only how to do mathematics, and how to apply mathematics to real world problems, but also how to become a selfless giver that takes care of the people around him, a leader that trust his team. I would like to thank Professor Zari Rachev for his support through the years of my doctoral research. Many thanks to Professor Robert Frey for having created such a unique graduate program. I also thank Dr.Riedel for having invented the TIB form 20 years ago along with Andrew, the work later becomes the foundation of my dissertation. I thank Professor Christopher Bishop, who played a key role in my introduction of complex analysis.

My gratitude also goes out to various of students and my colleagues: Xu Dong, Tengjie Jia, Xiaoping Zhou, Pengyuan Shao, Angela Cao, Ruoyu Zhou, Youxi Lin, Yu Mu, Xiang Shi, Mike Tiano, Tim De Lise, Tianyu Lu, Hua Mo, Riyu Yu, and Ke Zhang, etc. It's my great honor to study and work with you. I'm thankful to the staff in the deparment: Christine Rota and Laurie Dalessio. I would like to express my special appreciation to Jaehyung Choi and Edourd Coakley, who spent innumerable efforts on reviewing and revising my dissertation.

I would like to thank a group of my fridends named "finger laker": Yuyang Zhang, Alex Wang, Susie Sun, Nick Song, Mark Huang, Si Chen, Wei Cao, Ran Ma, Yiran Wang, Pu Zhao, Yujiao Chen and Feng Chen. We shared our happiness and sadness ever since the finger lake trip and gradually became family.

At last, I want to give my utmost thanks to my parents.

1 Introduction

1.1 Motivating Trends

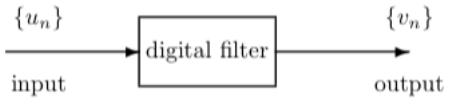
Linear system identification and reduction technique have been intensively studied for decades [1, 2, 5, 19, 38]. We are primarily interested in the discrete-time linear time invariant state-space dynamical systems. These systems can be used to model and predict time series.

Dynamical systems are the basic framework for modeling and control of an enormous variety of complex systems of scientific interest or industrial value. Examples include heat transfer, temperature control in various media, signal propagation and interference in electric circuits, wave propagation and vibration suppression in large structures, and behavior of micro-electro-mechanical systems. Direct numerical simulation of the associated models has been one of the few available means for studying complex underlying physical phenomena. However, the ever increasing need for improved accuracy requires the inclusion of ever more detail in the modeling stage, leading inevitably to ever larger-scale, ever more complex dynamical systems. Simulations in such large-scale settings often lead to unmanageably large demands on computational resources, which is the why we care about large scale problem.

It is often desirable to represent a high order system by a lower order system, the use of statistical, fundamental, hidden, or explicit factor models, and principal component analysis are common examples within quantitative finance applied fields. In most instances, such lower order models provide reasonable accuracy for realization, control, and computational purposes. The fast development and use of smaller processors, such as personal and minicomputers, in the design, analysis, and implementation of dynamic systems enhance the importance and increased interest in effective model reduction schemes. One of the main results of the dissertation is a model reduction algorithm that beats the commercial packages algorithms (Example 34).

We choose to focus on the state space realization because of their definite advantages. Such realizations are amenable to hardware implementations; they can be easily generalized to the time-variant case, and they act as useful utilities in the formulation and analysis of the given system. It is possible to apply the vast knowledge of matrix theory in the analysis, while the non-uniqueness of state space realizations provides the design engineer with the choice of using one that is better suited for the purpose at hand. This choice may be governed

Figure 1: Digital Filter



by truncation errors, roundoff errors, sensitivity issues, etc.

Let's assume the market as a digital filter, in reality, we only get to observe the output information, i.e. the quotes, trades, volumes, etc., which we will call such system identification problem the "Blind system identification" (BSI) [45]. BSI is a fundamental signal processing technology aimed at retrieving unknown information of a system from its output only. This technology is particularly suitable for applications where all the available data are generated from an unknown system driven by an unknown input. For example, returns modeled and predicted for one or more future period can be combined with optimization to perform a portfolio selection, such as described in the Appendix. The word "blind" simply means that the system's input is not available to (cannot be seen by) the signal processor. Note that if either the system function or the input signal is known, it becomes a more standard and simpler problem. The notion of BSI (or the like, such as blind deconvolution) has become well known since the early 1980s. During the 1990s, there has been an increasing research interest devoted to BSI. Unlike most of the work in the 1980s, research in the 1990s tended to explore to a higher degree the diversities inherent in multiple-output systems. The multiple-output systems arise from multisensor systems, multichannel data acquisition, or fractional sampling systems. Research articles recently produced by the signal processing community contain a significant amount of new knowledge that can be applied by many other communities, such as the seismic community, speech community, and medical community.

In the 1990s, Mullhaupt and Riedel developed the triangular input balanced (TIB) representation for single-input single-output (SISO) discrete-time linear state-space systems, which was proven to be a desirable parametrization for system identification and reduction [13]. The key feature of the TIB parametrization is it has an orthogonal basis which is uniquely and accurately determined by the poles of the system. The other benefit of TIB is that the feedback/ad-

vance matrix has a band ratio structure or consecutive subblock structure which enables fast updates of the state in the state space system. An extension to the Schur-Horn theorem relating to the consecutive subblock structure is provided in the Appendix. Naturally we want to expand the TIB SISO case to the multi-input multi-output (MIMO) case, which can deal with multi-dimension data instead of a single time series, e.g. different foreign exchange pairs, bid and ask, trade prices and trade volumes, etc. The motivation of the dissertation is thus: we want to find the MIMO TIB representation and related band ratio structure, as well as identification and reduction algorithms for MIMO TIB systems.

1.2 Problem Overview

A causal discrete-time linear time invariant systems can be charaterized by a transfer function, which is the z-transformation of an impulse response $\{h_n\}$ (see section 2),

$$H(z) = \sum_{n=0}^{\infty} h_n z^n. \tag{3}$$

Then let's consider a state space discrete-time linear time invariant d-dimensional model,

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t), & (4)\\ y(t) &= Cz(t) + Dx(t), & (5) \end{aligned}$$

where $y(t)$ is a sequence of d-dimensional measurement vectors and $z(t)$ is a state vector of dimension n. The impulse response corresponds to the state space system is given by

$$\hat{h}_k = CA^{k-1}B, k > 0 \tag{6}$$

and $\hat{h}_0 = D$. The reduction/realization problem is of determining a quadruple (A, B, C, D) such that $\left\{\hat{h}_n\right\}$ is close to $\{h_n\}$. There are several criteria to measure how close the estimated impulse reponse to the "real" impuse response $\{h_n\}$, and among them the H_2 criterion and H_∞ criterion are heavliy studied [24, 25, 26, 34, 50, 55]. The H_2 criterion minimizes the expression

$$\sum_{k=0}^{\infty} |h_k - \hat{h}_k|^2, \tag{7}$$

while the H_∞ criterion minimizes

$$\| \begin{pmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & & \\ h_3 & & \ddots & \\ \vdots & & & \ddots \end{pmatrix} - \begin{pmatrix} \hat{h}_1 & \hat{h}_2 & \hat{h}_3 & \cdots \\ \hat{h}_2 & \hat{h}_3 & & \\ \hat{h}_3 & & \ddots & \\ \vdots & & & \ddots \end{pmatrix} \|_2. \tag{8}$$

However, as we explain in section 2.7 we seek to miminize the information distance, which is the l_2 difference between the log-transfer function and log-estimated transfer function.

In the MIMO case, the first problem we solve is the parametrization, for which the work of Hanzon and Olivi [21, 24] provides insight. We then generalize the existing hybrid model reduction algorithm and adaptive model identification algorithms to the MIMO case. In addition, we generalized the band ratio and consecutive subblock structure of SISO TIB to matrix case.

1.3 Organization of The Dissertation

The paper is structured as follows: Section 2 introduces some background knowledge we need for our results; In Section 3, we discuss the SISO TIB case from several different angles, and also the relationship with the celebrated lattice filter; Section 4 studies the result of [22], which is a generalization of MIMO TIB representation and the lattice filter. We give the parametrization a new interpretation, and discuss the realationship between the lossless transfer functions, realization matrices and tangential Schur data. In Section 5, we find several structures for the MIMO TIB representation as consequences of the low grade attribute; in Section 6, efficient model reduction and identification algorithms for MIMO linear system in VMOA (Space of analytic functions of Vanishing Mean Oscillation) are developed by making use of FFT and TIB form.

2 Mathematical Preliminary

2.1 Digital Filters and Transfer Functions

A causal *digital filter* is a tranformation that takes any digital signal $\{x_n\}_{n=0}^{\infty} \subseteq \mathbb{R}^{p\times 1}$, called an *input* signal, to a digital signal $\{y_n\}_{n=0}^{\infty} \subseteq \mathbb{R}^{q\times 1}$, called the corresponding output signal. A digital filter is also characterized as convolution with a sequence of complex numbers (matrices),

$$h_0, h_1, h_2, \ldots$$

in the sense that that the output $\{y_n\}$ is obtained from the input $\{x_n\}$ by convolution with the "filter sequence" $\{h_n\}_{n=0}^{\infty} \subseteq \mathbb{R}^{p\times q}$ as

$$\{y_n\} = \{h_n\} * \{x_n\}, \tag{9}$$

or equivalently,

$$y_n = \sum_{i=0}^{\infty} h_i x_{n-i.} \tag{10}$$

As is clear from this definition, a digital filter satisfies three properties: linearity, time-invariance, and causality [1]. For this reason, the digital filter is also called a *causal linear time-invariant (LTI) system,* the sequence $\{h_n\}$used to define the system is also called the *impulse response.* A LTI system with impulse response $\{h_n\}$ is called *stable* if and only if $\{h_n\}$ lives in l^1, which is

$$\sum_{n=0}^{\infty} |h_n| < \infty. \tag{11}$$

When $\{h_n\}$ is a finite sequence (i.e., $h_n = 0, \forall n > M$, where M is some non-negative integer), the digital filter is called a *Finite Impulse Response (FIR)* digital filter. If infinitely many h_n are nonzero, the filter is called an *Infinite Impulse Response (IIR)* digital filter.

The *transfer function* of the filter $H(z)$ is the z-transformation of the impulse response $\{h_n\}$,

$$H(z) := \sum_{n=0}^{\infty} h_n z^n. \tag{12}$$

Since the z-transformation takes convolution of sequences to algebraic multipli-

cation of polynomials, we have

$$Y(z) = H(z) X(z), \tag{13}$$

where

$$X(z) := \sum_{n=0}^{\infty} x_n z^n, \; Y(z) := \sum_{n=0}^{\infty} y_n z^n, \tag{14}$$

that is, the spectrum of the output signal is obtained by multiplying the spectrum of the input signal by the transfer function. An IIR filter with a rational transfer function $H(z)$ is *stable* if and only if all the poles of the rational function $H(z)$ lie in the open unit disk $|z| < 1$ [1]. Our main research interest is the stable LTI system.

2.2 State Space Description

The state space system

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t), & (15) \\ y(t) &= Cz(t) + x(t), & (16) \end{aligned}$$

where A, B, C, D are $m \times m, m \times p, q \times m$, and $q \times p$ matrices independent of t, effects a convolution of $\{x_t\}$and $\{h_t\}$, where $\{h_t\}$ is the impulse response

$$h_k = CA^{k-1}B, k > 0, \tag{17}$$

and $h_0 = I$. In (15) the vector z_t which contains all of the important information is called the *state*, the matrix A which governs the state is called the *system/advance/feedback matrix*, and the matrix B that dictates the input the control sequence is called the *control/innovation matrix*. In equation (16) the matrix C that describes how the state is measured is called the *observation/measurement matrix*. Here $\{y_t\}$ and $\{x_t\}$ are $p \times 1$ and $q \times 1$ column vectors respectively, hence the system may also be called *Multi-Input/Multi-Output (MIMO)* system. In the special case when both p and q are equal to 1, it's called a *Single-Input/Single-Output (SISO)* system.

By the Cayley-Hamilton theorem, there is a polynomial $p(z)$ of degree less than or equal to m, such that $p(A) = 0$. We normally write this polynomial as

$$p(A) = \prod_{k=1}^{m} (\lambda_k I - A) \tag{18}$$

where $\{\lambda_k\}$ are the eigenvalues of A (replicated for multiplicity). In particular this allows us to express A^m as a linear combination of $I, A, \ldots, A^{m-1}$. In view of this, the (semi-infinite) Krylov matrix

$$\begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \tag{19}$$

has colums spanned by the columns of the *reachability matrix*

$$\mathcal{R} = \begin{pmatrix} B & AB & \cdots & A^{m-1}B \end{pmatrix}. \tag{20}$$

The Krylov matrix is full rank if and only if the reachability matrix $\mathcal{R}$ is non-singular. A state space system with this property is callled *reachable*. We also have the *reachability Grammian*

$$\mathcal{P} = \mathcal{R}\mathcal{R}^*. \tag{21}$$

The reachability Grammian is positive definite if and only if the reachability matrix is non-singular, i.e. the reachability Krylov matrix has full rank. Similarly, a state space system is called *observable* if and only if the *observability matrix*

$$\mathcal{O} = \begin{pmatrix} C \\ CA \\ \vdots \\ CA^{m-1} \end{pmatrix} \tag{22}$$

is non-singular, in which case the *observability Grammian*

$$\mathcal{Q} = \mathcal{O}^*\mathcal{O} \tag{23}$$

is positive definite. It follows that a state space system is *minimal* if and only if it is both reachable and observable. We now form the infinite block Hankel matrix

$$\Gamma_H = \begin{pmatrix} h_1 & h_2 & h_3 & \cdots \\ h_2 & h_3 & \cdots & \\ h_3 & \cdots & & \\ \cdots & & & \end{pmatrix}, \tag{24}$$

where $\{h_t\}$ is the impulse response. Any (block) Hankel matrix has a full rank factorization in Krylov matrices, which corresponds to a state space system, and

such a full rank factorization is called minimal. *Kronecker's Theorem* states that the infinite Hankel matrix Γ_H has finite rank if and only if the singular part

$$H_s(z) = \sum_{t=1}^{\infty} h_t z^t \tag{25}$$

is a (strictly) proper rational function in z-domain [1]. Furthermore, the rank of Γ_H agrees with the number of poles of $H_s(z)$, which multiplication is taken into consideration.

Under a change of coordinates of the state space from z to Tz, where T is any non-singular transformation,

$$\begin{aligned} Tz_{t+1} &= \left(TAT^{-1}\right)(Tz_t) + (TB)\,x_t, & (26)\\ y_t &= \left(CT^{-1}\right)(Tz_t) + x_t. & (27) \end{aligned}$$

The coordinate change corresponds to the change of the state space system parameters from (A, B, C) to $\left(TAT^{-1}, TB, CT^{-1}\right)$, while the impulse response (and Hankel matrix) is preserved:

$$\begin{aligned} \left(CT^{-1}\right)\left(TAT^{-1}\right)^k (TB) &= CT^{-1}TA^kT^{-1}TB & (28)\\ &= CA^kB. & (29) \end{aligned}$$

Observability, reachability and minimality are all preserved under the change of state space coordinates, as this transformation preserves the positive definiteness of the reachability Grammian and the observability Grammian, by Sylverster inertia.

From realizaiton theory it follows that any proper rational matrix-valued function $G(z)$ can be written in the form of

$$G(z) = D + C\left(zI_m - A\right)^{-1} B, \tag{30}$$

where (A, B, C, D) is an appropriate quadruple of matrices and m is the state space dimension [1]. The associated quadruple is called a *state space realization* of $G(z)$. To such a realization we associate the block-partitioned matrix

$$R = \begin{pmatrix} D & C \\ B & A \end{pmatrix}, \tag{31}$$

which we call the *realization matrix*.

2.3 Popapov Factorization and Blaschke-Potapov Factor

Let Σ be a $k \times k$ Hermitian unitary matrix . Note that Σ is unitarily similar to a signature matrix J, i.e., there exists a unitary $k \times k$ matrix U for which $U\Sigma U^*$ attains the form $\begin{pmatrix} I_q & \\ & -I_r \end{pmatrix}$ for some non-negative integers q and r with $q+r=k$. A square rational matrix function $\Theta(z)$ of size $k \times k$ is called *J-inner*, if at every point of analyticity of z of $\Theta(z)$ it satisfies

$$\Theta(z)^* J\Theta(z) \leq J, \quad |z| < 1, \tag{32}$$

$$\Theta(z)^* J\Theta(z) = J, \quad |z| = 1, \tag{33}$$

$$\Theta(z)^* J\Theta(z) \geq J, \quad |z| > 1. \tag{34}$$

The *McMillan degree* is the dimension of the feedback matrix when the system is minimal.

Theorem 1. *Every rational J-inner matrix-valued function (mvf) $U(\lambda)$ can be represented as a finite product of elementary Blaschke-Potapov factors. Moreover, if the McMillan degree of $U(\lambda)$ is equal to r, then $U(\lambda)$ may be expressed as the product of r primary Blaschke Potapov factors that are each normalized at a point α times a constant J-unitary matrix on either the left or the right.*

Proof. See Chapter 11 of Dym [7].

□

In [30], M.Olivi provides an alternative parametrization of the Blaschke-Potapov factors, namely,

$$\mathcal{B}_{\omega,u}(z) := I + \left(\frac{1-\bar{\omega}z}{z-\omega} - 1\right) uu^*, \tag{35}$$

with u an unit vector. It has simple properties such as

$$\det \mathcal{B}_{\omega,u}(z) = \frac{z-\omega}{1-\bar{\omega}z}, \tag{36}$$

and

$$\mathcal{B}_{\omega,u}(1/\bar{\omega})\, u = 0. \tag{37}$$

In the SISO case, we have the finite or infinite Blaschke product

$$B(z) = \prod_i \frac{z - \omega_i}{1 - \bar{\omega}_i z}. \tag{38}$$

The following lemma is well known.

Lemma 2. *The Blaschke product $B(z)$ is convergent, if and only if that*

$$\sum_i (1 - |\omega_i|) < \infty. \tag{39}$$

Proof. See [1].

□

2.4 Douglas-Shapiro-Shields Factorization

A strictly proper transfer function can be represented by means of the Douglas-Shapiro-Shields factorization

$$H = PG, \tag{40}$$

where G is rational lossless, P is rational unstable matrix, and G and H have same McMillan degree. The set of $\mathbb{C}^{p\times p}$-valued rational lossless functions of degree n will be denoted by $\mathcal{L}_n^p$. Note that the function H belongs to the set $\mathcal{H}(G)$, the orthogonal complement of H^2G into $\bar{H^2}$. Denotes the projection $\pi_n(G)$ of F onto $\mathcal{H}(G)$, in H^2 approximation case, since $\mathcal{H}(G)$ is a vector space, we want to minimize the objective function

$$\min_G \|F - \pi_n(G)\|_2^2. \tag{41}$$

It was proved that $\mathcal{L}_n^p$ is a smooth manifold, therefore we now deal with a minimization problem over a manifold, which is the nice set-up to use differential tools. In the scalar case, the manifold $\mathcal{L}_p^n$ is trivial, in particular, it is an open subset of a Euclidean space. In the multivariable case, the main difficulty was to find a nice parametrization for the manifold $\mathcal{L}_n^p$, which was done by means of Schur parameters, determined by a tangential Schur algorithm, and used as local coordinates[22].

2.5 Real Schur Decomposition and Lanczos Algorithm

The *Real Schur Form* will be used for getting the real form of MIMO TIB when there are complex conjugate pairs of poles, and the Lanczos algorithm will be used for computing the parital SVD of block Hankel matrices.

Theorem 3. *For any matrix $A \in \mathbb{C}^{n\times n}$, there exists a unitary matrix $Q \in \mathbb{C}^{n\times n}$ such that*

$$A = QTQ^*, \tag{42}$$

where T is lower triangular with the eigenvalues of A on its diagonal. Furthermore, Q can be chosen so that the eigenvalues appear in any order along the diagonal.

This is the Schur decomposition. When $A \in \mathbb{R}^{n\times n}$, A can have complex conjugate pair of eigenvalues which would lead to complex elements in Q and T. To avoid complex number in computations, we must lower our expectations and compromise with the calculation of an alternative decomposition known as the *real Schur decomposition.*

Theorem 4. *(Real Schur Decomposition) If $A \in \mathbb{R}^{n\times n}$, then there exists an orthogonal $Q \in \mathbb{R}^{n\times n}$ such that*

$$Q^T AQ = \begin{pmatrix} R_{11} & R_{12} & \cdots & R_{1m} \\ & R_{22} & \cdots & R_{2m} \\ & & \ddots & \vdots \\ & & & R_{mm} \end{pmatrix} \tag{43}$$

where each R_{ii} is either a 1×1 matrix or a 2×2 matrix having complex conjugate eigenvalues.

Proof. See Chapter 7 of [60].

□

The theorem shows that any real matrix is orthogonally similar to an upper (lower) quasi-triangular matrix.

The *Lanczos algorithm* is an iterative algorithm to compute the m eigenvalues and eigenvectors of an order n linear system with a limited number of operations $O(mn)$, where m is much smaller than n. To get he eigenvalues of matrix A, unlike the power method, the more advanced algorithms such as Arnoldi's algorithm and the Lanczos algorithm, save the information of a series of vectors $A^j v, j = 0, 1, \ldots, n-1$, and use the Gram-Schmidt process or Householder algorithm to reorthogonalize them into a baiss spanning the Krylov subspace corresponding to the matrix A. The Lanczos algorithm can be viewed as a simplified Arnoldi's algorithm in that it applies to Hermitian matrices. The m's step of the algorithm transforms the matrix A into a tridiagonal matrix T_{mm}. After the matrix T_{mm} is calculated, one can solve its eigenvalues $\lambda_i^{(m)}$ and their corresponding eigenvalues $u_i^{(m)}$ using the QR algorithm in as little as $O\left(m^2\right)$ work. It can be proved that the eigenvalues are approximate eigenvalues of the original matrix A. One thing worth pointing out is that to get the eigenvalues of A, we don't necessarily know A, all we need is the multiplication with A.

2.6 Low Grade Matrices and Consecutive Subblock Product

A notation of low grade was developed by A.Mullhaupt and K.Riedel in [3].

Definition 5. The upper (lower) grade of a matrix M, written $ugrade(M)$ ($lgrade(M)$) is the maximum rank of a part of symmetric partition above (below) the diagonal. The *grade* of a matrix M is the maximum rank of an off diagonal part of a symmetric partition, that is, $grade(M) = \max\{lgrade(M), ugrade(M)\}$.

Here are some examples of low grade commonly used matrices.

Proposition 6. *Companion matrices and Jordan matrices have grade* 1. *Elememtary row (column) operation matrices, Householder, and hyperbolic Householder transformations, Givens and signed Givens rotations all have grade* 1.

[3] also provides some algebraic properties of the grade and lgrade.

Theorem 7. *Let M_1 and M_2 be an $n \times n$ matrix. Then*

1. $lgrade\,(M_1 + M_2) \leq lgrade\,(M_1) + lgrade\,(M_2)$;
2. $lgrade\,(M_1 M_2) \leq lgrade\,(M_1) + lgrade\,(M_2)$;
3. $lgrade\left(M^{-1}\right) = lgrade\,(M)$;
4. $grade\left(M^{-1}\right) = grade\,(M)$.

Before further discussion of the relationship between low grade and band ratio structure, we first want to introduce the concept of consecutive subblock products, which is an equally important matrix structure widely used in matrix analysis, digital signal processing and control theory.

Definition 8. Let F_k be an $n \times n$ matrix such that $F_k e_j = e_j$ and $e_j^* F_k = e_j^*$for $j < k$ and for $j > k+d$. Then $M = F_1 F_2 \cdots F_{n-d}$ is called a *consecutive subblock product of order d.*

In other words, F_k is the identity except for a $(d+1) \times (d+1)$ block on the main diagonal,

$$F_k = \begin{pmatrix} I_{k-1} & & & & \\ & * & \cdots & * & \\ & \vdots & \ddots & \vdots & \\ & * & \cdots & * & \\ & & & & I_{n-k-d} \end{pmatrix}. \tag{44}$$

Mullhaupt and Riedel pointed out that consecutive subblock products are low grade matrices [3].

The following theorem discusses the connection between lgrade matrices and the existence of their band fraction representation. We will then prove any advance matrix in a TIB pair has low-grade, and therefore a band fraction representation.

Theorem 9. *Suppose $lgrade\,(M) \leq d$. Then there are matrices L and H such that $LM = H$ with L lower triangular and $lwidth\,(L) \leq d$ and $lwidth\,(H) \leq d$ and $uwidth\,(H) \leq uwidth\,(M)$.*

Now we want to restrain our interest to the application of low grade structure on TIB representations.

Theorem 10. *If M is unitary, then $lgrade(M) = ugrade(M)$.*

Proof. Since $M^{-1} = M^*$, then $lgrade(M) = ugrade\left(M^{-1}\right) = ugrade(M)$. □

Corollary 11. *Let M be unitary and Hessenberg, then $grade(M) = 1$.*

2.7 Information Geometry of Linear Systems

Definition 12. Given a family of probability distributions parametrized by θ, the *Fisher information matrix* is defined as

$$\begin{aligned} F(\theta) &= E\left(\left(\partial_\theta \log p(x,\theta)\right)\left(\partial_\theta \log p(x,\theta)\right)^T \mid \theta\right) && (45)\\ &= \int p(x,\theta)\left(\partial_\theta \log p(x,\theta)\right)\left(\partial_\theta \log p(x,\theta)\right)^T d\mu(x) && (46)\\ &= \int \left(\frac{\partial_\theta p(x,\theta)}{\sqrt{p(x,\theta)}}\right)\left(\frac{\partial_\theta p(x,\theta)}{\sqrt{p(x,\theta)}}\right)^T d\mu(x). && (47) \end{aligned}$$

Because it's positive definite, the Fisher information matrix determines an inner product as follows

$$\begin{aligned} \langle u, v\rangle_{F(\theta)} &:= u^T F(\theta) v, && (48)\\ \|u\|_{F(\theta)} &= \sqrt{\langle u, u\rangle_{F(\theta)}}. && (49) \end{aligned}$$

The *Jeffreys prior* is a non-informative prior distribution on parameter space that is propotional to the square root of the determinant of the Fisher information matrix.

The distance between two points P, Q is given by the shortest length of all piecewise smooth path γ_P^Q joining these two points. The length of a path $\gamma(t)$ is $\int_\gamma \|\gamma'(t)\|_{F(\theta)} dt$, a curve that encompasses this shortest path is geodesic.

Definition 13. Suppose that $\theta(t), 0 \leq t \leq 1$ is a smooth path connecting $\theta(0) = \theta_0$ and $\theta(1) = \theta_1$, the information distance between θ_0 and θ_1 $I(\theta_0 \mid \theta_1)$ is defined by

$$I(\theta_0 \| \theta_1) = \min \int_0^1 \sqrt{\dot{\theta}(t)^T F(\theta) \dot{\theta}(t)} dt. \tag{50}$$

The information distance is independent with respect to smooth changes of coordinates on the manifold.

We will use a univariate normal distribution parametrized by mean and standard deviation as an example. The PDF for univariate Gaussian distribution is

$$f(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(\frac{-(x-\mu)^2}{2\sigma^2}\right), \tag{51}$$

each point in the upper half plane $H = \left\{(\mu, \sigma) \in \mathbb{R}^2 \mid \sigma > 0\right\}$ associates a distribution in univariate normal family. A proper distance arises from the Fisher information matrix, which would be computed as follows:

$$\frac{\partial \log f(x, \mu, \sigma)}{\partial \mu} = \frac{(x-\mu)}{\sigma^2}, \tag{52}$$

$$\frac{\partial \log f(x, \mu, \sigma)}{\partial \sigma} = -\frac{1}{\sigma} + \frac{(x-\mu)^2}{\sigma^3}, \tag{53}$$

then the Fisher information matrix is

$$F(\mu, \sigma) = \begin{pmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{2}{\sigma^2} \end{pmatrix}, \tag{54}$$

so the expression for the metric is

$$ds_F^2 = \frac{d\mu^2 + 2d\sigma^2}{\sigma^2}. \tag{55}$$

Amari and Nagaoka [31] pointed out in order to analyze the similarity between two systems or two time series, and consider the problems of approximation, estimation, and dimension lowering, it is necessary to consider the space consisting of all such systems and analyze its geometric structure. A discrete linear

system is described by a transfer function

$$H(z) = \sum_{i=0}^{\infty} h_i z^i, \tag{56}$$

with

$$x_t = H(z)\,\varepsilon_t. \tag{57}$$

We assume $\sum_{i=0}^{\infty} |h_i| < \infty$ so it is a stable system. If the input ε_t is i.i.d. white Gaussian noise, then the output is a stationary Gaussian time series. Amari and Nagaoka [31] studied the space of Gaussian time series L which consists of the set of all S that satisfy

$$\int_{-\pi}^{\pi} |\log S(\omega)|^2 d\omega < \infty. \tag{58}$$

The metric tensor of the statistical manifold of $\begin{pmatrix} \cdots & x_{-1} & x_0 & x_1 & \cdots \end{pmatrix}$ can be derived from the spectral density function:

$$g_{ij}(\xi) = \frac{1}{2\pi} \int_{-\pi}^{\pi} (\partial_i \log S)(\partial_j \log S)\, d\omega, \tag{59}$$

where the partial derivative is the derivative with respect to the coordinate system of the model parameter ξ. In z-domain, we have

$$g_{ij} = \frac{1}{2\pi i} \int_{|z|=1} (\partial_i \log H(z))(\partial_j \log H(z))^* \frac{dz}{z}. \tag{60}$$

From Choi and Mullhaupt [61], for a transfer function parametrized in the form

$$\log H(z) = \sum_{i=0}^{\infty} a_i z^i, \tag{61}$$

the Fisher information matrix would be the identity in cepstrum coordinates, thus the information distance from white noise in terms of the power series of the logarithm of the transfer function is the norm of the Hardy space of the disc $H^2(\mathbb{D})$. Information geometry is an outcome of the investigation of the differential geometric structure on manifolds of probablity distributions, with the Riemannian metric defined by Fisher information matrix.

3 Triangular Input Balanced Form (SISO case)

3.1 The Representation

The TIB representation of systems was introduced by A.Mullhaupt and K.Riedel [13, 14]. Consider a state space system with innovation form

$$\begin{aligned} z_{t+1} &= Az_t + Bx_t, && (62)\\ y_t &= Cz_t + x_t, && (63) \end{aligned}$$

the TIB (Triangular Input Balanced Form) means A is lower triangular and (A, B) is a input balanced pair, i.e.

$$AA^* + BB^* = I. \tag{64}$$

In section 2.2, we mentioned for any non-singular transformation T, the coordinate change

$$\begin{aligned} A &\rightarrow TAT^{-1}, && (65)\\ B &\rightarrow TB, && (66)\\ C &\rightarrow CT^{-1}, && (67) \end{aligned}$$

preserves impulse response. We claim that starting from any state space realization of a LTI system, we are able to find the equivalent TIB representation. Let P be the grammian matrix of the semi-infinite Krylov matrix,

$$\begin{aligned} P &= \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ B^*A^{2*} \\ \vdots \end{pmatrix} && (68)\\ &= \sum_{k=0}^{\infty} BA^kA^{k*}B^*, && (69) \end{aligned}$$

then we have the Stein equation,

$$APA^* + BB^* = P, \tag{70}$$

where P is positive definite. The Schur decomposition of A is

$$A = QSQ^*, \tag{71}$$

where Q is unitary and S is lower triangular. Thus,

$$S\left(Q^*PQ\right)S^* + Q^*BB^*Q = Q^*PQ, \tag{72}$$

where Q^*PQ is also positive definite. We then take the Cholesky decomposition of $Q^*PQ = LL^*$,

$$SLL^*S^* + Q^*BB^*Q = LL^* \tag{73}$$

which can also be written as

$$\left(L^{-1}SL\right)\left(L^{-1}SL\right)^* + \left(L^{-1}Q^*B\right)\left(L^{-1}Q^*B\right)^* = I. \tag{74}$$

Therefore, if we choose the coordinate transformation T as $T = L^{-1}Q^*$, in which case

$$\tilde{A} = L^{-1}SL, \tag{75}$$

$$\tilde{B} = L^{-1}Q^*B, \tag{76}$$

such that $\left(\tilde{A}, \tilde{B}\right)$ is an input balanced pair and $\tilde{A}$ is a product of three lower triangular matrices which is lower triangular. The poles of the linear system are the eigenvalues of A, notice that they are preserved by coordinate change. When A is triangular, the eigenvalues of A are just the diagonal elements. Actually, in the TIB case, given the poles of the system the TIB pair is uniquely determined. Here is heuristic explanation, consider the partial isometry,

$$\begin{pmatrix} B & A \end{pmatrix} = \begin{pmatrix} b_1 & a_{11} & & & \\ b_2 & a_{21} & a_{22} & & \\ \vdots & \vdots & & \ddots & \\ b_k & a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix}. \tag{77}$$

The rows are perpendicular to each other, thus where are $\frac{k(k+1)}{2}$ equations constraining $\frac{k(k+1)}{2} + k$ unknown elements. If we know the poles (k elements on the diagonal), then the remaining elements are determined.

Also in the input balanced case, we know something about the singular values of A. Suppose that B is a $k \times 1$ vector, then there are $k-1$ vectors $v_1, v_2, \ldots, v_{k-1}$ that are perpendicular to B, so

$$AA^* v_j + BB^* v_j = AA^* v_j = v_j, \tag{78}$$

for $j = 1, 2, \ldots, k-1$, and

$$AA^* B + BB^* B = B \tag{79}$$

$$\Rightarrow \quad AA^* B = (1 - B^* B)\, B. \tag{80}$$

Therefore, A has singular values

$$\sigma_1(A) = \sigma_2(A) = \cdots = \sigma_{k-1}(A) = 1, \tag{81}$$

and

$$\sigma_k(A) = 1 - B^* B. \tag{82}$$

Immediately, we have

$$||Au|| < ||u||, \forall u \tag{83}$$

which says A is a contraction.

For a real rational transfer function, it's possible that the system has complex conjugate pairs of poles. To avoid a complex representation in TIB form, , we could rotate the feedback matrix A to make it real. Consider the following decomposition,

$$\begin{pmatrix} A_1 & & & \\ * & A_2 & & \\ \vdots & \ddots & \ddots & \\ * & \cdots & * & A_n \end{pmatrix} = \begin{pmatrix} L_1 & & & \\ * & L_2 & & \\ \vdots & \ddots & \ddots & \\ * & \cdots & * & L_n \end{pmatrix} \begin{pmatrix} Q_1 & & & \\ * & Q_2 & & \\ \vdots & \ddots & \ddots & \\ * & \cdots & * & Q_n \end{pmatrix} \tag{84}$$

where A_k are 2×2 matrices with real elements and conjugate pairs of eigenvalues, Q_k are 2×2 unitary matrices also with real elements, and L_k are 2×2 lower triangular matrices with real elements. Let L_k have the magnitude of the poles $|\lambda_k|$ on its diagonal, the properties of Meixner functions in [14] indicates that

the only choice for other elements are $\left(1-|\lambda_k|^2\right)$, or actually $\left(|\lambda_k|^2-1\right)$. Thus

$$A_k = \begin{pmatrix} |\lambda_k| & \\ 1-|\lambda_k|^2 & |\lambda_k| \end{pmatrix} Q_k. \tag{85}$$

When

$$Q_k = \begin{pmatrix} \cos\theta_k & \sin\theta_k \\ -\sin\theta_k & \cos\theta_k \end{pmatrix}, \tag{86}$$

since post-multiplying an unitary matrix preserves the determinant, we would only need to consider the trace,

$$\begin{aligned} & tr\left(L_k Q_k\right) && (87)\\ = \;& 2|\lambda_k|\cos\theta_k + \sin\theta_k - |\lambda_k|^2 \sin\theta_k && (88)\\ = \;& 2\left(|\lambda_k|\cos\frac{\theta_k}{2} + \sin\frac{\theta_k}{2}\right)\left(\cos\frac{\theta_k}{2} - |\lambda_k|\sin\frac{\theta_k}{2}\right). && (89) \end{aligned}$$

Now let's apply the transformation

$$|\lambda_k| = \tan\frac{\phi_k}{2}, \tag{90}$$

then

$$\begin{aligned} & \frac{|\lambda_k|}{\sqrt{1+|\lambda_k|^2}}\cos\frac{\theta_k}{2} + \frac{1}{\sqrt{1+|\lambda_k|^2}}\sin\frac{\theta_k}{2} && (91)\\ = \;& \sin\frac{\phi_k}{2}\cos\frac{\theta_k}{2} + \cos\frac{\phi_k}{2}\sin\frac{\theta_k}{2} && (92)\\ = \;& \sin\frac{\phi_k+\theta_k}{2}, && (93) \end{aligned}$$

and

$$\begin{aligned} & \frac{1}{\sqrt{1+|\lambda_k|^2}}\cos\frac{\theta_k}{2} - \frac{|\lambda_k|}{\sqrt{1+|\lambda_k|^2}}\sin\frac{\theta_k}{2} && (94)\\ = \;& \cos\frac{\phi_k}{2}\cos\frac{\theta_k}{2} - \sin\frac{\phi_k}{2}\sin\frac{\theta_k}{2} && (95)\\ = \;& \cos\frac{\phi_k+\theta_k}{2}, && (96) \end{aligned}$$

therefore

$$
\begin{aligned}
tr\left(L_k Q_k\right) &= 2\left(1+|\lambda_k|^2\right)\sin\frac{\phi_k+\theta_k}{2}\cos\frac{\phi_k+\theta_k}{2} && (97)\\
&= \left(1+|\lambda_k|^2\right)\sin\left(\phi_k+\theta_k\right) && (98)\\
&= 2Re\left(\lambda_k\right) && (99)
\end{aligned}
$$

gives us the angle

$$
\theta_k = \arccos\left(\frac{2Re\left(\lambda_k\right)}{1+|\lambda_k|^2}\right) - 2\arctan\left(|\lambda_k|\right), \tag{100}
$$

which determines the transformation from complex form to real form.

3.2 Orthogonal Function Point of View

[19] provides an interesting brief history of ideas of rational orthonomal bases related to system identification. It says:

> The first mention of rational orthonormal bases seems to have occurred in 1925-1928 with the independent work of Takenaka and Malmquist. The particular bases considered were those that will be denoted as 'generalized orthonormal basis functions'.

Mullhaupt and Riedel [12] gave credit to Ninness and Gustafsson, however Ninness-Gustaffson bases and Takenaka-Malmquist functions are the same thing. Given the prior knowledge of the poles $\{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ of the discrete-time linear system, Ninness and Gustafsson[18] derived a set of orthonormal bases functions in $H_2(\mathbb{T})$:

$$
\mathcal{B}_n(q) = \left(\frac{\sqrt{1-|\lambda_n|^2}}{q-\lambda_n}\right)\prod_{k=0}^{n-1}\left(\frac{1-\lambda_k^* q}{q-\lambda_k}\right) \tag{101}
$$

satisfies

$$
\langle \mathcal{B}_n, \mathcal{B}_m \rangle = \frac{1}{2\pi i}\int_{\mathbb{T}} \mathcal{B}_n(z)\overline{\mathcal{B}_m(z)}\frac{dz}{z} = \delta_{mn}. \tag{102}
$$

The bases can be obtained from $\left\{\frac{1}{q-\lambda_1}, \frac{1}{q-\lambda_2}, \ldots, \frac{1}{q-\lambda_n}\right\}$ by Gram-Schmidt orthogonalization. If we pick poles at 0, we will get FIR filter, the one pole and two poles case correspond to *Laguerre* and *Kautz* filters. It is also worth pointing out that the bases are closely related to the *HAMBO trasformation* [19], which

maps the transfer function from z-domain to λ-domain in such a way that the poles of the transformed systems are faster (closer to the origin) than those for the original system while stability and orthogonality are preserved.

3.3 Schur Algorithm Point of View and Lattice Filter

The lattice filter is a well-known recursive filter structure. Most of the properties find their origins in the fist half of the 20th centrury, thanks to the prominent work done by Schur and Szego, i.e. Schur recursion and Szego polynomial. Regalia [5] is good reference for tapped state lattice filter. The idea of lattice filter is to write the rational transfer function $H(z)$ as

$$H(z) = \sum_{k=0}^{M} \nu_k \frac{\hat{D}_k(z)}{D_M(z)} \tag{103}$$

where $\hat{D}_k(z)$ are polynomials of degree $k-1$.

Starting from the denominator of $H(z)$, a natural idea is using a similar polynomial $D_k(z)$ which has the same coefficients in reverse order, i.e.

$$\hat{D}_k(z) = z^k D_k\left(z^{-1}\right). \tag{104}$$

For $k = M, M-1, \ldots, 1$, if $D_M(z)$ is a minimum phase polynomial, which means that all roots lie strictly outside the unit circle, then Rouche's theorem indicates that all $D_k(z)$ are minimum phase, therefore $|\frac{\hat{D}_k(0)}{D_k(0)}| < 1$. Suppose $D_M(z)$ is minimum phase, then we can set

$$\sin\theta_k = \frac{\hat{D}_k(0)}{D_k(0)} \tag{105}$$

and take positive $\cos\theta_k$. Also we denote $s_k = \sin\theta_k$ and $c_k = \cos\theta_k$, $\hat{D}_{k-1}(z)$ and $D_{k-1}(z)$ of the degree $k-2$ can be obtained from $\hat{D}_k(z)$ and $D_k(z)$ by Schur recursion,

$$\begin{pmatrix} D_{k-1}(z) \\ z\hat{D}_{k-1}(z) \end{pmatrix} = \frac{1}{c_k} \begin{pmatrix} 1 & -s_k \\ -s_k & 1 \end{pmatrix} \begin{pmatrix} D_k(z) \\ \hat{D}_k(z) \end{pmatrix}, \tag{106}$$

which can be rearranged in the form

$$\begin{pmatrix} D_{k-1}(z) \\ \hat{D}_k(z) \end{pmatrix} = \begin{pmatrix} -s_k & c_k \\ c_k & s_k \end{pmatrix} \begin{pmatrix} z\hat{D}_{k-1}(z) \\ D_k(z) \end{pmatrix}. \tag{107}$$

Recall the state space representation,

$$\begin{pmatrix} z_{t+1} \\ w_{t+1} \end{pmatrix} = Q \begin{pmatrix} z_t \\ x_t \end{pmatrix} \tag{108}$$

$$y_t = \begin{pmatrix} \nu_0 & \nu_1 & \cdots & \nu_M \end{pmatrix} \begin{pmatrix} z_t \\ w_t \end{pmatrix}, \tag{109}$$

where

$$Q = \begin{pmatrix} A & B \\ * & * \end{pmatrix}. \tag{110}$$

Immediately we have

$$\begin{pmatrix} \frac{\hat{D}_0(z)}{D_M(z)} \\ \frac{\hat{D}_1(z)}{D_M(z)} \\ \vdots \\ \frac{\hat{D}_M(z)}{D_M(z)} \end{pmatrix} x_t = Q \begin{pmatrix} z\frac{\hat{D}_0(z)}{D_M(z)} \\ z\frac{\hat{D}_1(z)}{D_M(z)} \\ \vdots \\ z\frac{\hat{D}_{M-1}(z)}{D_M(z)} \\ 1 \end{pmatrix} x_t, \tag{111}$$

which is

$$\begin{pmatrix} \hat{D}_0(z) \\ \hat{D}_1(z) \\ \vdots \\ \hat{D}_M(z) \end{pmatrix} = Q \begin{pmatrix} z\hat{D}_0(z) \\ z\hat{D}_1(z) \\ \vdots \\ z\hat{D}_{M-1}(z) \\ D_M(z) \end{pmatrix}. \tag{112}$$

Denote

$$Q_k = \begin{pmatrix} I_{k-1} & & & \\ & -s_k & c_k & \\ & c_k & s_k & \\ & & & I_{M-k} \end{pmatrix}, \tag{113}$$

and from the rearranged Schur recursion,

$$Q_1 Q_2 \cdots Q_M \begin{pmatrix} z\hat{D}_0(z) \\ z\hat{D}_1(z) \\ \vdots \\ z\hat{D}_{M-1}(z) \\ D_M(z) \end{pmatrix} \tag{114}$$

$$= Q_1 Q_2 \cdots Q_{M-1} \begin{pmatrix} z\hat{D}_0(z) \\ z\hat{D}_1(z) \\ \vdots \\ D_{M-1}(z) \\ \hat{D}_M(z) \end{pmatrix} \tag{115}$$

$$= \cdots \tag{116}$$

$$= Q_1 Q_2 \cdots Q_k \begin{pmatrix} z\hat{D}_0(z) \\ \vdots \\ z\hat{D}_k(z) \\ D_{k-1}(z) \\ \hat{D}_k(z) \\ \vdots \\ \hat{D}_M(z) \end{pmatrix} \tag{117}$$

$$= \cdots \tag{118}$$

$$= \begin{pmatrix} \hat{D}_0(z) \\ \hat{D}_1(z) \\ \vdots \\ \hat{D}_M(z) \end{pmatrix}. \tag{119}$$

Finally we know Q can be represented as the product of the sequence of Q_k,

$$Q = Q_1 Q_2 \cdots Q_M, \tag{120}$$

which is a consecutive subblock product with $c_1, c_2, \ldots, c_M$ on the subdiagonal. Because each Q_i is unitary, Q is unitary. If we partition

$$Q = \begin{pmatrix} A & B \\ g & \nu_0 \end{pmatrix} \tag{121}$$

where A is the $M \times M$ feedback matrix of the lattice filter, then

$$AA^* + BB^* = I_M \tag{122}$$

which indicates that the lattice filter has input balanced form. The other thing we want to verify is that

$$\det(I - zA) = c_0 D_M(z), \tag{123}$$

where c_0 is some constant. Denote

$$P_k = \prod_{i=1}^{k} \begin{pmatrix} I_{i-1} & & & \\ & -s_i & c_i & \\ & c_i & s_i & \\ & & & I_{k-i} \end{pmatrix} \tag{124}$$

and

$$A_k = P_{k-1} \begin{pmatrix} I_{k-1} & \\ & -s_k \end{pmatrix}. \tag{125}$$

Then we have the recursive updating formula,

$$P_k = \begin{pmatrix} A_k & c_k P_{k-1} e_k \\ c_k e_k^T & s_k \end{pmatrix} \tag{126}$$

and

$$A_{k+1} = \begin{pmatrix} A_k & -s_{k+1} c_k P_{k-1} e_k \\ c_k e_k^T & -s_k s_{k+1} \end{pmatrix}. \tag{127}$$

We will prove the following by induction.

Proposition 14. *Using the notation above, we have*

$$\det(zI - A_k) = \left(1 / \prod_{j=k+1}^{M} c_j \right) \hat{D}_k(z) \tag{128}$$

and

$$\det\left(\begin{pmatrix} zI_k & \\ & 0 \end{pmatrix} - P_k \right) = -\left(1 / \prod_{j=k+1}^{M} c_j \right) D_k(z). \tag{129}$$

Proof. When $k = 1$,

$$A_1 = -s_1, \tag{130}$$

$$P_1 = \begin{pmatrix} -s_1 & c_1 \\ c_1 & s_1 \end{pmatrix}, \tag{131}$$

and

$$\det(z - A_1) = z + s_1, \tag{132}$$

$$\det \begin{pmatrix} z + s_1 & -c_1 \\ -c_1 & -s_1 \end{pmatrix} = -(s_1 z + 1), \tag{133}$$

gives us the initial condition of induction. At every step of the induction, we have two stages,

$$\det(zI_{k+1} - A_{k+1}) \tag{134}$$

$$= \det \begin{pmatrix} zI_k - A_k & s_{k+1} c_k P_{k-1} e_k \\ -c_k e_k^T & z + s_k s_{k+1} \end{pmatrix} \tag{135}$$

$$= \det \begin{pmatrix} zI_k - A_k & s_{k+1} c_k P_{k-1} e_k \\ 0 & z \end{pmatrix} \tag{136}$$

$$+ \det \begin{pmatrix} zI_k - A_k & s_{k+1} c_k P_{k-1} e_k \\ -c_k e_k^T & s_k s_{k+1} \end{pmatrix} \tag{137}$$

$$= z \det(zI_k - A_k) - s_{k+1} \det \left(\begin{pmatrix} zI_k & \\ & 0 \end{pmatrix} - P_k \right) \tag{138}$$

$$= \left(1 / \prod_{j=k+1}^{M} c_j \right) \left(z \hat{D}_k(z) + s_{k+1} D_k(z) \right) \tag{139}$$

$$= \left(1 / \prod_{j=k+1}^{M} c_j \right) c_{k+1} \hat{D}_{k+1}(z) \tag{140}$$

$$= \left(1 / \prod_{j=k+2}^{M} c_j \right) \hat{D}_{k+1}(z), \tag{141}$$

and

$$\det\left(\begin{pmatrix} zI_{k+1} & \\ & 0 \end{pmatrix} - P_{k+1}\right) \tag{142}$$

$$= \det\begin{pmatrix} zI_{k+1} - A_{k+1} & c_{k+1}P_k e_{k+1} \\ -c_{k+1}e_{k+1}^T & -s_{k+1} \end{pmatrix} \tag{143}$$

$$= s_{k+1}\det\left(zI_{k+1} - A_{k+1}\right) - c_{k+1}^2 \det\left(\begin{pmatrix} zI_k & \\ & 0 \end{pmatrix} - P_k\right) \tag{144}$$

$$= s_{k+1}\left(1/\prod_{j=k+2}^{M} c_j\right)\hat{D}_{k+1}(z) + c_{k+1}^2\left(1/\prod_{j=k+1}^{M} c_j\right)D_k(z) \tag{145}$$

$$= -\left(1/\prod_{j=k+2}^{M} c_j\right)D_{k+1}(z) \tag{146}$$

$$= -\left(1/\prod_{j=k+2}^{M} c_j\right)D_{k+1}(z) \tag{147}$$

□

Starting from a TIB system, we can obtain the lattice representation by writing down the rational transfer function first. It is also possible to go from TIB to lattice form directly. The *bulge chasing* technique can be used for this purpose, suppose we have a unitary matrix

$$U = \begin{pmatrix} A & b \\ g & \delta \end{pmatrix} \tag{148}$$

where A is upper triangular, so U has the form

$$U = \begin{pmatrix} * & \cdots & * & * \\ & \ddots & \vdots & \vdots \\ & & * & * \\ * & \cdots & * & * \end{pmatrix}. \tag{149}$$

We want to apply a similarity transformation to U to make it take the form

$$Q = \begin{pmatrix} * & \cdots & * & * \\ * & \cdots & * & * \\ & \ddots & \vdots & \vdots \\ & & * & * \end{pmatrix}. \tag{150}$$

Then the upper left part of Q is the desired feedback matrix of a lattice system. To better demostrate our "bulge chasing" algorithm, we show a 4×4 example here. Now

$$U = \begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ * & * & * & * & * \end{pmatrix}, \tag{151}$$

let's rotate the first two rows and columns to introduce a zero on the last row from the bottom,

$$\begin{pmatrix} * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ * & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & * & * & * & * \end{pmatrix}. \tag{152}$$

Then rotate the second and third rows and columns to introduce another zero on the last row,

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & * & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & * & * & * \end{pmatrix}. \tag{153}$$

The next step is to rotate the first two rows and columns again to introduce a

zero on the third row,

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & * & * & * \end{pmatrix}. \tag{154}$$

Then it's the turn of the third and fourth rows and columns to introduce another zero on the last row,

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \\ 0 & 0 & * & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}. \tag{155}$$

And the second and third rows and columns,

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}. \tag{156}$$

At last, let's rotate the first and second rows and columns,

$$\begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix} \rightarrow \begin{pmatrix} * & * & * & * & * \\ * & * & * & * & * \\ 0 & * & * & * & * \\ 0 & 0 & * & * & * \\ 0 & 0 & 0 & * & * \end{pmatrix}. \tag{157}$$

Thus the lattice form are obtained by a sequence of rotaions based from TIB.

4 MIMO TIB Form

4.1 Hanzon-Olivi-Peeters parametrization

For every pair of matrices $U, V \in \mathbb{C}^{(p+1)\times(p+1)}$, we associate a map acting on a proper rational $p \times p$ matrix function $G(z)$ as follows:

$$\mathcal{F}_{U,V} : G(z) \to F_1(z) + \frac{F_2(z) F_3(z)}{z - F_4(z)}, \tag{158}$$

with $F_1(z)$ of size $p \times p$, $F_2(z)$ of size $p \times 1$, $F_3(z)$ of size $1 \times p$ and $F_4(z)$ a scalar, where each is specified by the partitioning of

$$V \begin{pmatrix} 1 & \\ & G(z) \end{pmatrix} U^* = F(z) \tag{159}$$

$$= \begin{pmatrix} F_1(z) & F_2(z) \\ F_3(z) & F_4(z) \end{pmatrix}. \tag{160}$$

Proposition 16 builds up a direct connection between the state space realization of $G(z)$ and $\mathcal{F}_{U,V}(G(z))$[22]. We give a simpler, alternative proof of Proposition 16, using the Crabtree-Haynsworth quotient formula [32] :

Lemma 15. *The composition of Schur complement is the same as the direct Schur complement for the same partition.*

Proof. Suppose we have a matrix parition where J and $\begin{pmatrix} E & F \\ H & J \end{pmatrix}$ are invertible,

$$X = \begin{pmatrix} A & B & C \\ D & E & F \\ G & H & J \end{pmatrix} \tag{161}$$

then the larger Schur complement of X is

$$Y = \begin{pmatrix} A & B \\ D & E \end{pmatrix} - \begin{pmatrix} C \\ F \end{pmatrix} J^{-1} \begin{pmatrix} G & H \end{pmatrix} \tag{162}$$

$$= \begin{pmatrix} A - CJ^{-1}G & B - CJ^{-1}H \\ D - FJ^{-1}G & E - FJ^{-1}H \end{pmatrix}, \tag{163}$$

its complement is

$$Y_c = A - CJ^{-1}G - \left(B - CJ^{-1}H\right)\left(E - FJ^{-1}H\right)^{-1}\left(D - FJ^{-1}G\right), \quad (164)$$

whereas the smaller Schur complement of X is

$$\begin{aligned}
& Z && (165)\\
= \; & A - \begin{pmatrix} B & C \end{pmatrix} \begin{pmatrix} E & F \\ H & J \end{pmatrix}^{-1} \begin{pmatrix} D \\ G \end{pmatrix} && (166)\\
= \; & A - \begin{pmatrix} B & C \end{pmatrix} \begin{pmatrix} I & 0 \\ -J^{-1}H & I \end{pmatrix} \begin{pmatrix} \left(E - FJ^{-1}H\right)^{-1} & 0 \\ 0 & J^{-1} \end{pmatrix} \begin{pmatrix} I & -FJ^{-1} \\ 0 & I \end{pmatrix} \begin{pmatrix} D \\ G \end{pmatrix} && (167)\\
= \; & A - \begin{pmatrix} B - CJ^{-1}H & C \end{pmatrix} \begin{pmatrix} \left(E - FJ^{-1}H\right)^{-1} & 0 \\ 0 & J^{-1} \end{pmatrix} \begin{pmatrix} D - FJ^{-1}G \\ G \end{pmatrix} && (168)\\
= \; & A - CJ^{-1}G - \left(B - CJ^{-1}H\right)\left(E - FJ^{-1}H\right)^{-1}\left(D - FJ^{-1}G\right) && (169)\\
= \; & Y_c && (170)
\end{aligned}$$

□

Proposition 16. *Let $G(z)$ be a proper rational transfer function and (U, V) be a pair of $(p+1) \times (p+1)$ matrices, then*

$$\tilde{G}(z) = \mathcal{F}_{U,V}(G(z)) \quad (171)$$

is well-defined. Let (A, B, C, D) be a state space realization of $G(z)$ with n-dimensional space. Then a state space realization $\left(\tilde{A}, \tilde{B}, \tilde{C}, \tilde{D}\right)$ of $\tilde{G}(z)$ with $(n+1)$-dimensional state space is given by:

$$\begin{pmatrix} \tilde{D} & \tilde{C} \\ \tilde{B} & \tilde{A} \end{pmatrix} = \begin{pmatrix} V & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & D & C \\ & B & A \end{pmatrix} \begin{pmatrix} U^* & \\ & I_n \end{pmatrix}. \quad (172)$$

Proof. First since $G(z)$ is proper, $F_4(z)$ is proper as well and therefore $z - F_4(z)$ does not vanish identically so that $\mathcal{F}_{U,V}(G(z))$ is well-defined. Now observe the

realization of $F(z)$ is given by

$$\begin{pmatrix} D_F & C_F \\ B_F & A_F \end{pmatrix} = \begin{pmatrix} V & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & D & C \\ & B & A \end{pmatrix} \begin{pmatrix} U^* & \\ & I_n \end{pmatrix} \tag{173}$$

According to lemma 15, the $p \times p$ Schur complement of $\begin{pmatrix} D_F & C_F \\ B_F & A_F \end{pmatrix} - \begin{pmatrix} 0 & 0 \\ 0 & zI_{n+1} \end{pmatrix}$ is the $p \times p$ Schur complement of the $(p+1) \times (p+1)$ Schur complement $F(z) - \begin{pmatrix} 0 & 0 \\ 0 & z \end{pmatrix}$ which is $F_1(z) + \frac{F_2(z)F_3(z)}{z - F_4(z)}$.

□

The main result that connects state space realizations and the tangential Schur algorithm can now be stated as follows.

Theorem 17. *Let G be a $p \times p$-all-pass function of degree n. Such a function admits a balanced realization*

$$G(z) = D + C(zI_n - A)^{-1}B \tag{174}$$

such that the associated realization matrix

$$R = \begin{pmatrix} D & C \\ B & A \end{pmatrix} \tag{175}$$

is unitary. Let R_{n-1} be a $(n+p) \times (n+p)$-unitary realization matrix of $G_{n-1}(z)$, then a unitary realization matrix R_n of $G(z)$ is given by

$$R_n = \begin{pmatrix} V_n & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & \\ & R_{n-1} \end{pmatrix} \begin{pmatrix} U_n^* & \\ & I_n \end{pmatrix}, \tag{176}$$

where U and V are unitary $(p+1) \times (p+1)$ complex matrices depending on u, v of size p and scalar w as follows

$$U = \begin{pmatrix} \xi u & I_p - (1 + w\eta)uu^* \\ \bar{w}\eta & \xi u^* \end{pmatrix}, \tag{177}$$

$$V = \begin{pmatrix} \xi v & I_p - (1 - \eta)\frac{vv^*}{||v||^2} \\ \eta & -\xi v^* \end{pmatrix}, \tag{178}$$

with

$$\xi = \frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2||v||^2}}, \tag{179}$$

$$\eta = \frac{\sqrt{1-||v||^2}}{\sqrt{1-|w|^2||v||^2}}. \tag{180}$$

The tangential Schur algorithm consists of repeating this process, thus providing a sequence of all-pass functions $G_k(z)$ of degree k, satisfying the interpolation condition

$$G_k(1/\bar{w_k})u_k = v_k, ||v_k|| < 1. \tag{181}$$

Proof. See [22].

□

4.2 Unified Framework

In this subsection, the recursive matrix factorization of Olivi, Hanzon and Peeters [22] of all-pass transfer functions is interpreted as a unified framework for orthogonal filters, including the well-known lattice filters, and as we determine here, TIB filters of both SISO and MIMO cases. We will first address several observations and remarks on the generic form and then restrict it to some special cases by choosing different w and v.

From equation (176),

$$R_n = \Phi \begin{pmatrix} I_{n+1} & \\ & R_0 \end{pmatrix} \Psi, \tag{182}$$

where

$$\Phi = \begin{pmatrix} V_n & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & V_{n-1} & \\ & & I_{n-1} \end{pmatrix} \cdots \begin{pmatrix} I_{n-2} & & \\ & V_2 & \\ & & I_2 \end{pmatrix} \begin{pmatrix} I_{n-1} & & \\ & V_1 & \\ & & 1 \end{pmatrix}, \tag{183}$$

$$\Psi = \begin{pmatrix} I_{n-1} & & \\ & U_1^* & \\ & & 1 \end{pmatrix} \begin{pmatrix} I_{n-2} & & \\ & U_2^* & \\ & & I_2 \end{pmatrix} \cdots \begin{pmatrix} 1 & & \\ & U_{n-1}^* & \\ & & I_{n-1} \end{pmatrix} \begin{pmatrix} U_n^* & \\ & I_n \end{pmatrix} \tag{184}$$

Here R_0 is unitary, Φ and Ψ are unitary consecutive subblock products [4],

moreover Φ is upper p-Hessenberg while Ψ is lower p-Hessenberg.

Now we will discuss several special cases:

Case 1. SISO, $w = 0$;

Case 2. SISO, $v = 0$;

Case 3. MIMO, $w = 0$;

Case 4. MIMO, $v = 0$.

In Case 1, since u has unit length, without loss of generality we can assume $u = 1$, then we have $\xi = 1$ and $\eta = \sqrt{1 - |v|^2}$. Therefore

$$U = I_2 \tag{185}$$

and

$$V = \begin{pmatrix} v & \sqrt{1-|v|^2} \\ \sqrt{1-|v|^2} & -\bar{v} \end{pmatrix}. \tag{186}$$

In addition Ψ is the identity matrix and Φ is a consecutive subblock product of elementary unitary matrices. Also we know that R is a unitary upper Hessenberg matrix with $\left\{\sqrt{1-|v_1|^2}, \sqrt{1-|v_2|^2}, \ldots, \sqrt{1-|v_n|^2}\right\}$ on the subdiagonal. Then A is upper Hessenberg, immediately we recognize that this is the lattice filter form with the reflection coefficients $\left\{\sqrt{1-|v_1|^2}, \sqrt{1-|v_2|^2}, \ldots, \sqrt{1-|v_n|^2}\right\}$. In other words, lattice filter form can be obtained by scalar tangential Schur interpolation interpolated at 0.

In Case 2, since $G(1/\bar{w})u = 0, |u| = 1$, we know $G(1/\bar{w})$ has to be zero which implies that w is the pole of the all-pass function. Now $\xi = \sqrt{1-|w|^2}$ and $\eta = 1$, then

$$U = \begin{pmatrix} \sqrt{1-|w|^2}u & -w \\ \bar{w} & \sqrt{1-|w|^2}\bar{u} \end{pmatrix}, \tag{187}$$

and

$$V = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \tag{188}$$

which gives

$$\Phi = \begin{pmatrix} & 1 & \\ I_n & & \\ & & 1 \end{pmatrix}. \tag{189}$$

Suppose

$$R_0 = \begin{pmatrix} D_0 & C_0 \\ B_0 & A_0 \end{pmatrix} \tag{190}$$

and let's denote U_0 such that

$$U_0^* = \begin{pmatrix} B_0 & A_0 \\ D_0 & C_0 \end{pmatrix} \tag{191}$$

accordingly, then

$$\begin{aligned} \begin{pmatrix} B & A \\ D & C \end{pmatrix} &= \begin{pmatrix} & I_{n+1} \\ 1 & \end{pmatrix} \begin{pmatrix} B & A \\ D & C \end{pmatrix} && (192) \\ &= \left[\begin{pmatrix} I_n & & \\ & & 1 \\ & 1 & \end{pmatrix} \Phi^{-1} \right] \left[\Phi \begin{pmatrix} I_n & \\ & R_0 \end{pmatrix} \Psi \right] && (193) \\ &= \begin{pmatrix} I_n & \\ & U_0^* \end{pmatrix} \Psi && (194) \end{aligned}$$

which is a unitary lower Hessenberg matrix with superdiagonal $\{w_1, w_2, \ldots, w_n\}$. So A is lower triangular and (A, B) is a TIB pair. The TIB form is the scalar tangential Schur interpolation at the poles of the system.

Case 3 seems to be MIMO lattice filter, though we are not able to find related literature yet.

We claim Case 4 is the MIMO TIB form. When the Schur vectors v are zero vectors, we know that the interpolation points w have to be poles and we call the interpolation vectors u the "null vectors". In this case, $\xi = \sqrt{1 - |w|^2}$ and $\eta = 1$, again we have

$$U = \begin{pmatrix} \sqrt{1 - |w|^2} u & I_p - (1 + w)\, uu^* \\ \bar{w} & \sqrt{1 - |w|^2} \bar{u} \end{pmatrix},$$

and

$$V = \begin{pmatrix} & I_p \\ 1 & \end{pmatrix}. \tag{195}$$

Similar to the SISO TIB case,

$$\Phi = \begin{pmatrix} & I_p & \\ I_n & & \\ & & 1 \end{pmatrix}, \tag{196}$$

suppose

$$R_0 = \begin{pmatrix} D_0 & C_0 \\ B_0 & A_0 \end{pmatrix} \tag{197}$$

and let's denote U_0 such that

$$U_0^* = \begin{pmatrix} B_0 & A_0 \\ D_0 & C_0 \end{pmatrix} \tag{198}$$

accordingly, then

$$\begin{pmatrix} B & A \\ D & C \end{pmatrix} = \begin{pmatrix} & I_{n+1} \\ I_p & \end{pmatrix} \begin{pmatrix} B & A \\ D & C \end{pmatrix} \tag{199}$$

$$= \left[\begin{pmatrix} I_n & & \\ & & 1 \\ & I_p & \end{pmatrix} \Phi^{-1} \right] \left[\Phi \begin{pmatrix} I_n & \\ & R_0 \end{pmatrix} \Psi \right] \tag{200}$$

$$= \begin{pmatrix} I_n & \\ & U_0^* \end{pmatrix} \Psi \tag{201}$$

which is an unitary lower p-Hessenberg matrix with p-superdiagonal $\{w_1, w_2, \ldots, w_n\}$. So A is lower triangular and (A, B) is a TIB pair. The MIMO TIB form is the matricial tangential Schur interpolation at the poles of the system. The TIB pair (A, B) can be parametrized by poles and "null vectors".

4.3 Relation to Potapov factorization

It was shown by Potapov (section (2.3)) that any lossless matrix valued function of McMillan degree n can be decomposed into a product of n Blaschke-Potapov factors of McMillan degree 1, where the Blaschke-Potapov factor is defined as

$$\mathcal{B}_{\omega,u}(z) = I + \left(\frac{1 - \bar{\omega} z}{z - \omega} - 1 \right) uu^*, \tag{202}$$

with u a unit vector [13] such that

$$\mathcal{B}_{\omega,u}\left(1/\bar{\omega}\right)u=0. \tag{203}$$

Therefore, given a set of poles $(\omega_1,\omega_2,\ldots,\omega_n)$ and null vectors $(u_1,u_2,\ldots,u_n)$, the product of the Blaschke-Potapov factor $\mathcal{B}_{\omega_i,u_i}$,

$$\mathcal{G}_k\left(z\right)=\mathcal{B}_{\omega_1,u_1}\left(z\right)\mathcal{B}_{\omega_2,u_2}\left(z\right)\cdots\mathcal{B}_{\omega_k,u_k}\left(z\right) \tag{204}$$

is lossless, and it satisfies the interpolation condition,

$$\mathcal{G}_k\left(1/\bar{\omega}_k\right)u_k=0. \tag{205}$$

From Case 4 in Section 4.2, we know the parameters of the Blaschke-Potapov factors are exact the tangential Schur data, in particular, the poles and null vectors in the TIB case, therefore, the lossless balanced transfer function can be obtained by poles and null vectors. In addition, any lossless balanced transfer function has a TIB realization.

4.4 From the MIMO TIB pair to tangential Schur data

Hanzon-Olivi-Peeters (HOP) recursion provides a way of obtaining the realization matrix from the tangential Schur data. Given the tangential Schur data, we are always able to find a coordinate change such that the rotated system has TIB form. In this subsection, we provide the map from the TIB form to poles and null vectors, which can be used to construct the transfer function corresponding to the original tangential Schur data.

Let's suppose

$$R_n=\begin{pmatrix}D_n & C_n\\ B_n & A_n\end{pmatrix} \tag{206}$$

and

$$R_{n-1}=\begin{pmatrix}D_{n-1} & C_{n-1}\\ B_{n-1} & A_{n-1}\end{pmatrix} \tag{207}$$

are the realization matrices of lossless tranfer function satisfies the tangential

Schur algorithm in the TIB sense, then recursively,

$$\begin{pmatrix} D_n & C_n \\ B_n & A_n \end{pmatrix} \tag{208}$$

$$= \begin{pmatrix} & I_p & \\ 1 & & \\ & & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & D_{n-1} & C_{n-1} \\ & B_{n-1} & A_{n-1} \end{pmatrix} \tag{209}$$

$$\times \begin{pmatrix} \sqrt{1-|\omega_n|^2}u_n^* & \omega \\ I_p - (1+\bar{\omega}_n)\, u_n u_n^* & \sqrt{1-|\omega|^2}u & \\ & & I_n \end{pmatrix} \tag{210}$$

$$= \begin{pmatrix} & D_{n-1} & C_{n-1} \\ 1 & & \\ & B_{n-1} & A_{n-1} \end{pmatrix} \begin{pmatrix} \sqrt{1-|\omega_n|^2}u_n^* & \omega_n \\ I_p - (1+\bar{\omega}_n)\, u_n u_n^* & \sqrt{1-|\omega_n|^2}u_n & \\ & & I_n \end{pmatrix} \tag{211}$$

$$= \begin{pmatrix} D_{n-1} - (1+\bar{\omega}_n)\, D_{n-1}u_n u_n^* & \sqrt{1-|\omega_n|^2}D_{n-1}u_n & C_{n-1} \\ \sqrt{1-|\omega_n|^2}u_n^* & \omega_n & \\ B_{n-1}\,(I_p - (1+\bar{\omega}_n)\, u_n u_n^*) & \sqrt{1-|\omega_n|^2}B_{n-1}u_n & A_{n-1} \end{pmatrix} \tag{212}$$

which leads to the updating formulae,

$$D_n = D_{n-1} - (1+\bar{\omega}_n)\, D_{n-1}u_n u_n^* \tag{213}$$

$$C_n = \begin{pmatrix} \sqrt{1-|\omega_n|^2}D_{n-1}u_n & C_{n-1} \end{pmatrix} \tag{214}$$

$$B_n = \begin{pmatrix} \sqrt{1-|\omega_n|^2}u_n^* \\ B_{n-1}\,(I_p - (1+\bar{\omega}_n)\, u_n u_n^*) \end{pmatrix} \tag{215}$$

$$A_n = \begin{pmatrix} \omega_n & \\ \sqrt{1-|\omega_n|^2}B_{n-1}u_n & A_{n-1} \end{pmatrix} \tag{216}$$

Since the realization matrix R_n is unitary, we have $AA^* + BB^* = I$, and (A, B) is a input balanced pair. Moreover equation (216) indicates that A is triangular.

We observed that under the TIB form, the tangential Schur interpolation data (poles and null vectors) can be obtained from the realization matrix.

Theorem 18. *Given a TIB pair (A, B) from the realization matrix of a lossless system of McMillan degree n, the poles $\{\omega_1, \ldots, \omega_n\}$ and null vectors $\{u_1, \ldots, u_n\}$*

can be recursively computed as follows:

$$\begin{aligned} \omega_{n+1-i} &= A_{ii}, && (217)\\ u_{n+1-i} &= \frac{B^*_{1,1:end}}{\|B_{1,1:end}\|}, && (218)\\ B &= B_{2:end,1:end} - \left(1 + \frac{1}{\bar{w}_{n+1-i}}\right) B_{2:end,1;end} u_{n+1-i} u^*_{n+1-i} && (219) \end{aligned}$$

Proof. It is a direct computation from recursive formula (216).

□

Theorem 18 is an 'inverse' version of the Hanzon-Olivi-Peeters formula which allows us to construct efficient algorithms for model identification and reduction in Section 6. By 'inverse' we mean that the Hanzon-Olivi-Peeters formula gives the realization matrix of a lossless transfer function from the tangential Schur interpolation data, while Theorem 18 computes the tangential Schur interpolation data from the realization matrix.

5 Matrix Structures of the MIMO TIB form

5.1 Consecutive Subblock Product Structure

We studied the structures such as the consecutive subblock product and the band fraction of the MIMO TIB form which are the consequences of the low grade feature of TIB.

Theorem 19. *If (A, B) is a TIB pair with lower triangular $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times p}$, then A has low-grade p.*

Proof. see [3].

□

From equation (176),

$$R_n = \Phi \begin{pmatrix} I_{n+1} & \\ & R_0 \end{pmatrix} \Psi, \qquad (220)$$

where

$$\Phi = \begin{pmatrix} V_n & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & V_{n-1} & \\ & & I_{n-1} \end{pmatrix} \cdots \begin{pmatrix} I_{n-2} & & \\ & V_2 & \\ & & I_2 \end{pmatrix} \begin{pmatrix} I_{n-1} & & \\ & V_1 & \\ & & 1 \end{pmatrix}, \quad (221)$$

$$\Psi = \begin{pmatrix} I_{n-1} & & \\ & U_1^* & \\ & & 1 \end{pmatrix} \begin{pmatrix} I_{n-2} & & \\ & U_2^* & \\ & & I_2 \end{pmatrix} \cdots \begin{pmatrix} 1 & & \\ & U_{n-1}^* & \\ & & I_{n-1} \end{pmatrix} \begin{pmatrix} U_n^* & \\ & I_n \end{pmatrix} \quad (222)$$

We already know that Φ and Ψ are unitary consecutive subblock products [4], and Φ is upper p-Hessenberg while Ψ is lower p-Hessenberg.

Proposition 20. *lwidth* $\Phi \leq p$, *ugrade* $\Phi \leq p$; *uwidth* $\Psi \leq p$, *lgrade* $\Psi \leq p$.

Proof. It's a direct consequence of theorem 4.2 in [4].

□

Moreover, the inverse of Ψ can be written as

$$\Psi^{-1} = \Psi^* \quad (223)$$

$$= \begin{pmatrix} U_n & \\ & I_n \end{pmatrix} \begin{pmatrix} 1 & & \\ & U_{n-1} & \\ & & I_{n-1} \end{pmatrix} \cdots \begin{pmatrix} I_{n-2} & & \\ & U_2 & \\ & & I_2 \end{pmatrix} \begin{pmatrix} I_{n-1} & & \\ & U_1 & \\ & & 1 \end{pmatrix} \quad (224)$$

This expression has the same form as Φ, then R_n can be recognized as a ratio of two consecutive subblock products that have the same direction.

5.2 Band Fraction and Hessenberg Unitary Matrices

In this subsection, we study the band fraction structrue of Hessenberg unitray matrices, which derives the band fraction form for both the lattice filter and the TIB filter in the SISO case.

For a sequence of complex numbers, which have modulus smaller than 1,

$\rho = \{\rho_0, \rho_1, \ldots\}$, define band matrices pair $(M(\rho), N(\rho))$:

$$M = \begin{pmatrix} c_1 & & & & \\ s_1^* & c_2 & & & \\ & s_2^* & c_3 & & \\ & & \ddots & \ddots & \\ & & & s_{n-1}^* & c_n \end{pmatrix}, \tag{225}$$

$$N = \begin{pmatrix} s_1 & & & & \\ c_1 & s_2 & & & \\ & c_2 & s_3 & & \\ & & \ddots & \ddots & \\ & & & c_{n-1} & s_n \end{pmatrix}, \tag{226}$$

where $c_k = \frac{1}{\sqrt{1-|\rho_k|^2}}, s_k = \frac{\rho_k}{\sqrt{1-|\rho_k|^2}}$. First we know that for a nonzero sequence ρ, $A = (M(\rho))^{-1} N(\rho)$ and $B = (M(\rho))^{-1} e_1$ is a TIB pair.

Lemma 21 and Proposition 22 are results of Mullhaupt and Riedel [12]. We start from Lemma 21 and derive more band fraction structures for lattice.

Lemma 21. *For any nonzero sequence ρ inside the unit disk, by adding zeros to both front and end, then getting the corresponding band matrices $M(0, \rho, 0)$ and $N(0, \rho, 0)$, the band ratio $(M(0, \rho, 0))^{-1} N(0, \rho, 0)$ has the form of $\begin{pmatrix} 0 & 0 \\ U & 0 \end{pmatrix}$, where U is lower unitary Hessenberg matrix with ρ on superdiagonal.*

Proof.

$$(M(0, \rho, 0))^{-1} N(0, \rho, 0) \tag{227}$$

$$= \begin{pmatrix} 1 & \\ & M(\rho, 0) \end{pmatrix}^{-1} \begin{pmatrix} 0 & \\ e_1 & N(\rho, 0) \end{pmatrix} \tag{228}$$

$$= \begin{pmatrix} 1 & \\ & M(\rho, 0)^{-1} \end{pmatrix} \begin{pmatrix} 0 & \\ e_1 & N(\rho, 0) \end{pmatrix} \tag{229}$$

$$= \begin{pmatrix} 0 & \\ M(\rho, 0)^{-1} e_1 & M(\rho, 0)^{-1} N(\rho, 0) \end{pmatrix} \tag{230}$$

also

$$
\begin{aligned}
& (M(\rho,0))^{-1} N(\rho,0) && (231)\\
= \; & \begin{pmatrix} M(\rho) & \\ s_n e_n^T & 1 \end{pmatrix}^{-1} \begin{pmatrix} N(\rho) & \\ c_n e_n^T & 0 \end{pmatrix} && (232)\\
= \; & \begin{pmatrix} M(\rho)^{-1} & \\ -s_n e_n^T M(\rho)^{-1} & 1 \end{pmatrix} \begin{pmatrix} N(\rho) & \\ c_n e_n^T & 0 \end{pmatrix} && (233)\\
= \; & \begin{pmatrix} M(\rho)^{-1} N(\rho) & \\ & 0 \end{pmatrix}. && (234)
\end{aligned}
$$

From equation (230) and equation (234) we see $(M(0,\rho,0))^{-1} N(0,\rho,0)$ has the form $\begin{bmatrix} 0 & 0 \\ U & 0 \end{bmatrix}$. Since $M(\rho)^{-1}N(\rho)$ is lower triangular with ρ on diagonal, U is Henssenberg with ρ on superdiagonal. $\left(M(\rho,0)^{-1}e_1, M(\rho,0)^{-1}N(\rho,0)\right)$ is a TIB pair, with $M(\rho,0)^{-1}N(\rho,0)$ which has a zero last column, thus U is unitary. □

Proposition 22. *U defined above can be represented as a ratio of a lower band matrix and an upper band matrix with bandwidth* 2.

Proof. Actually,

$$U = M(\rho,0)^{-1} \begin{pmatrix} e_1 & N(\rho) \\ & c_n e_n^T \end{pmatrix} \tag{235}$$

is the band ratio representation.

□

Remark 23. If we write down the full version of the band matrices

$$P = M(\rho, 0) \tag{236}$$

$$= \begin{pmatrix} c_1 & & & & \\ s_1^* & c_2 & & & \\ & s_2^* & \ddots & & \\ & & \ddots & c_n & \\ & & & s_n^* & 1 \end{pmatrix}, \tag{237}$$

$$Q = \begin{pmatrix} e_1 & N(\rho) \\ & c_n e_n^T \end{pmatrix} \tag{238}$$

$$= \begin{pmatrix} 1 & s_1 & & & \\ & c_1 & s_2 & & \\ & & c_2 & \ddots & \\ & & & \ddots & s_n \\ & & & & c_n \end{pmatrix}, \tag{239}$$

Q can be obtained by taking Hermitian of P then shifting the diagonal. Also $PP^* = QQ^*$, thus $A = P^{-1}Q = P^*Q^{-*}$ can also be represented as product of an upper band matrix and the inverse of a lower band matrix.

In state space representations, a unitary realization matrix corresponds to an all-pass filter. If we partition U as $U = \begin{pmatrix} B & A \\ D & C \end{pmatrix}$, this is the TIB representation. On the other hand, a partition $U = \begin{pmatrix} D & C \\ B & A \end{pmatrix}$ is related to the lattice filter representation. In this case, A is lower Hessenberg with reflection coeffients $\kappa = \{\kappa_1, \kappa_2, \ldots, \kappa_n\}$, also $C = ce_1$. We claim that A also has a band ratio structure.

Proposition 24. *If A is the advance matrix from lattice filter with reflection coefficients $\kappa = \{\kappa_1, \kappa_2, \ldots, \kappa_n\}$, then A can be written as $A = P^{-1}Q$, where P is a lower band matrix and Q is an upper band matrix with bandwidth 2.*

Proof. By (22),

$$
\begin{aligned}
A &= \left(M(c,\kappa,0)^{-1} \begin{pmatrix} e_1 & N(c,\kappa) \\ & c_n e_{n+1}^T \end{pmatrix} \right)_{2:n+2,2:n+2} && (240)\\
&= \left(\begin{pmatrix} \frac{1}{\sqrt{1-|c|^2}} & \\ \frac{c^*}{\sqrt{1-|c|^2}} e_1 & M(\kappa,0) \end{pmatrix}^{-1} \right)_{2:n+2,:} \begin{bmatrix} N(c,\kappa) \\ c_n e_{n+1}^T \end{bmatrix} && (241)\\
&= \begin{pmatrix} -c^* M(\kappa,0)^{-1} e_1 & M(\kappa,0)^{-1} \end{pmatrix} \begin{pmatrix} \frac{c}{\sqrt{1-|c|^2}} & \\ \frac{1}{\sqrt{1-|c|^2}} e_1 & N(\kappa) \\ & \frac{1}{\sqrt{1-|\kappa_n|^2}} e_n^T \end{pmatrix} && (242)\\
&= M(\kappa,0)^{-1} \left(-\frac{|c|^2}{\sqrt{1-|c|^2}} e_1 e_1^T + \begin{pmatrix} \frac{1}{\sqrt{1-|c|^2}} e_1 & N(\kappa) \\ & \frac{1}{\sqrt{1-|\kappa_n|^2}} e_n^T \end{pmatrix} \right) && (243)\\
&= M(\kappa,0)^{-1} \begin{pmatrix} \sqrt{1-|c|^2} e_1 & N(\kappa) \\ & \frac{1}{\sqrt{1-|\kappa_n|^2}} e_n^T \end{pmatrix} && (244)
\end{aligned}
$$

Let

$$P = M(\kappa,0) \tag{245}$$

and

$$Q = \begin{pmatrix} \sqrt{1-|c|^2} e_1 & N(\kappa) \\ & \frac{1}{\sqrt{1-|\kappa_n|^2}} e_n^T \end{pmatrix}, \tag{246}$$

and obeserve that P is a lower band matrix and Q is an upper band matrix with bandwidth 2.

□

Remark 25. There is a free parameter c here, A is unitary if and only if $c = 0$. Also

$$QQ^* + |c|^2 e_1 e_1^T = PP^* \tag{247}$$

With the band ratio representation of the advance matrix, many computations related to the lattice filter can be accelerated. Now let us take a look at the reachability matrix $\mathcal{R} = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix}$ of the lattice filter with

reflection coefficient $\kappa = \{\kappa_1, \kappa_2, \ldots, \kappa_n\}$, from equation (24),

$$\mathcal{R}_{:,k+1} = A\mathcal{R}_{:,k} \tag{248}$$

$$\Rightarrow \quad P\mathcal{R}_{:,k+1} = Q\mathcal{R}_{:,k} \tag{249}$$

Defining $\kappa_0 = c$ and $\kappa_{n+1} = 0$. For $1 \leq m \leq n+1$, we have

$$\frac{\mathcal{R}_{m,k}}{\sqrt{1-|\kappa_{m-1}|^2}} + \frac{\mathcal{R}_{m+1,k}\kappa_m^*}{\sqrt{1-|\kappa_m|^2}} = \frac{\mathcal{R}_{m-1,k+1}\kappa_{m-1}}{\sqrt{1-|\kappa_{m-1}|^2}} + \frac{\mathcal{R}_{m,k+1}}{\sqrt{1-|\kappa_m|}}. \tag{250}$$

We denote corresponding orthonormal bases denote

$$f_j(z) = \sum_{i=0}^{\infty} \mathcal{R}_{j,i} z^i, j = 1, 2, \ldots, n. \tag{251}$$

Thus the recursive formula, equation (250) above is equivalent to

$$\frac{zf_m(z)}{\sqrt{1-|\kappa_{m-1}|^2}} + \frac{\kappa_m^* z f_{m+1}(z)}{\sqrt{1-|\kappa_m|^2}} = \frac{\kappa_{m-1} f_{m-1}(z)}{\sqrt{1-|\kappa_{m-1}|^2}} + \frac{f_m(z)}{\sqrt{1-|\kappa_m|^2}}, \tag{252}$$

which leads to the three term recursive formula

$$f_{m+1}(z) = \left(\frac{1}{z\kappa_m^*} - \frac{\sqrt{1-|\kappa_m|^2}}{\kappa_m^* \sqrt{1-|\kappa_{m-1}|^2}} \right) f_m(z) + \frac{\kappa_{m-1}\sqrt{1-|\kappa_m|^2}}{z\kappa_m^* \sqrt{1-|\kappa_{m-1}|^2}} f_{m-1}(z). \tag{253}$$

On the other hand, the lattice/Schur bases come with the natural recursive formula

$$f_{m+1}(z) = \frac{f_m(z) - f_m(0)}{z\left(1 - \overline{f_m(0)} f_m(z)\right)} \tag{254}$$

where $f_m(0) = \sqrt{1-|\kappa_m|^2}$.

5.3 Band Fraction Stucture of MIMO TIB form

Mullhaupt and Riedel [3] pointed out there are band fraction representation of TIB form for both SISO and MIMO case, in this section, we first introduce a novel algebraic expression for TIB pair and then explicitly construct a band ratio representation of TIB pair in terms of the poles and null vectors. In the algebraic expression we propose, the poles and null vectors appear quite separately, and we use it to derive the band fraction formula as follows:

For every TIB pair (A, B) where $A \in \mathbb{R}^{n\times n}$ and $B \in \mathbb{R}^{n\times p}(n > p)$, we will constructively prove that under the satisfaction of some singularity conditions, there exist two lower triangular band matrices $M, N \in \mathbb{R}^{n\times n}$ with bandwidth p, such that

$$A = M^{-1}N, \tag{255}$$

and

$$B = M^{-1}\begin{pmatrix} U \\ 0 \end{pmatrix} \tag{256}$$

for some unitary U.

In the state updating procedure, there is a multiplication of A, which costs $O\left(n^2\right)$ flops if A is triangular. With the band ratio representation of A, the multiplication can be accelerated to $O\left(np\right)$ flops. The SISO case of the band fraction representation has been discussed in [3], where the authors also point out the representation may be derived from the Takenaka-Malmquist functions. We extend the results to the multivariate case.

Theorem 26. *Suppose (A, B) is the TIB pair constructed from poles ω and null vectors u, and let*

$$\Omega = tril(\begin{pmatrix} u_n^* \\ u_{n-1}^* \\ \vdots \\ u_0^* \end{pmatrix}\begin{pmatrix} u_n & u_{n-1} & \cdots & u_0 \end{pmatrix}), \tag{257}$$

$$D^{(1)} = \begin{pmatrix} \frac{\bar{\omega}_n}{1+\bar{\omega}_n} & & \\ & \ddots & \\ & & \frac{\bar{\omega}_0}{1+\bar{\omega}_0} \end{pmatrix}, \tag{258}$$

$$D^{(2)} = \begin{pmatrix} \frac{|1+\omega_n|^2}{1-|\omega_n|^2} & & \\ & \ddots & \\ & & \frac{|1+\omega_0|^2}{1-|\omega_0|^2} \end{pmatrix}, \tag{259}$$

$$D^{(3)} = \begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & & \\ & \ddots & \\ & & \frac{\sqrt{1-|\omega_0|^2}}{1+\bar{\omega}_0} \end{pmatrix}, \tag{260}$$

then

$$A = D^{(3)}\left(D^{(2)} - \left(\Omega - D^{(1)}\right)^{-1}\right) D^{(3)}, \tag{261}$$

$$B = D^{(3)}\left(\Omega - D^{(1)}\right)^{-1} \begin{pmatrix} u_n^* \\ u_{n-1}^* \\ \vdots \\ u_0^* \end{pmatrix}. \tag{262}$$

Proof. First we observe

$$\frac{\sqrt{1-|\omega_0|^2}}{1+\bar{\omega}_0}\left(\frac{|1+\omega_0|^2}{1-|\omega_0|^2} - \left(1 - \frac{\bar{\omega}_0}{1+\bar{\omega}_0}\right)^{-1}\right)\frac{\sqrt{1-|\omega_0|^2}}{1+\bar{\omega}_0} \tag{263}$$

$$= \frac{1-|\omega_0|^2}{(1+\bar{\omega}_0)^2}\left(\frac{|1+\omega_0|^2}{1-|\omega_0|^2} - (1+\bar{\omega}_0)\right) \tag{264}$$

$$= \frac{1+\omega_0}{1+\bar{\omega}_0} - \frac{1-|\omega_0|^2}{1+\bar{\omega}_0} \tag{265}$$

$$= \omega_0 \tag{266}$$

$$= A_0, \tag{267}$$

and

$$\frac{\sqrt{1-|\omega_0|^2}}{1+\bar{\omega}_0}\left(1 - \frac{\bar{\omega}_0}{1+\bar{\omega}_0}\right)^{-1} u_0^* \tag{268}$$

$$= \sqrt{1-|\omega_0|^2}u_0^* \tag{269}$$

$$= B_0. \tag{270}$$

Now assume that A_{n-1} and B_{n-1} satisfy the formulae above, denote

$$\mu_n = \begin{pmatrix} u_n^* \\ u_{n-1}^* \\ \vdots \\ u_0^* \end{pmatrix}, \tag{271}$$

then

$$
\begin{aligned}
& D_n^{(3)} \left(\Omega_n - D_n^{(1)}\right)^{-1} \mu_n && (272)\\
= & \begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)} \end{pmatrix} \begin{pmatrix} 1 - \frac{\bar{\omega}_n}{1+\bar{\omega}_n} & \\ \mu_{n-1} u_n & \Omega_{n-1} - D_{n-1}^{(1)} \end{pmatrix}^{-1} \begin{pmatrix} u_n^* \\ \mu_{n-1} \end{pmatrix} && (273)\\
= & \begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)} \end{pmatrix} && (274)\\
& \times \begin{pmatrix} 1+\bar{\omega}_n & \\ -\left(1+\bar{\omega}_n\right) \left(\Omega_{n-1} - D_{n-1}^{(1)}\right)^{-1} \mu_{n-1} u_n & \left(\Omega_{n-1} - D_{n-1}^{(1)}\right)^{-1} \end{pmatrix} && (275)\\
& \times \begin{pmatrix} u_n^* \\ \mu_{n-1} \end{pmatrix} && (276)\\
= & \begin{pmatrix} \sqrt{1-|\omega_n|^2} & \\ -\left(1+\bar{\omega}_n\right) B_{n-1} u_n & D_{n-1}^{(3)} \left(\Omega_{n-1} - D_{n-1}^{(1)}\right)^{-1} \end{pmatrix} \begin{pmatrix} u_n^* \\ \mu_{n-1} \end{pmatrix} && (277)\\
= & \begin{pmatrix} \sqrt{1-|\omega_n|^2} u_n^* \\ B_{n-1} \left(I_p - \left(1+\bar{\omega}_n\right) u_n u_n^*\right) \end{pmatrix}, && (278)
\end{aligned}
$$

and B_n satisfies the updating formula. In addition,

$$
\begin{aligned}
& D_n^{(3)}\left(D_n^{(2)}-\left(\Omega_n-D_n^{(1)}\right)^{-1}\right)D_n^{(3)} && (279)\\
= \;& D_n^{(3)}D_n^{(2)}D_n^{(3)}-D_n^{(3)}\left(\Omega_n-D_n^{(1)}\right)^{-1}D_n^{(3)} && (280)\\
= \;& \begin{pmatrix} \frac{1+\omega_n}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}D_{n-1}^{(2)}D_{n-1}^{(3)}\end{pmatrix}-\begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}\end{pmatrix} && (281)\\
& \times\begin{pmatrix} 1+\bar{\omega}_n & \\ -\left(1+\bar{\omega}_n\right)\left(\Omega_{n-1}-D_{n-1}^{(1)}\right)^{-1}\mu_{n-1}u_n & \left(\Omega_{n-1}-D_{n-1}^{(1)}\right)^{-1}\end{pmatrix} && (282)\\
& \times\begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}\end{pmatrix} && (283)\\
= \;& \begin{pmatrix} \frac{1+\omega_n}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}D_{n-1}^{(2)}D_{n-1}^{(3)}\end{pmatrix} && (284)\\
& -\begin{pmatrix} \sqrt{1-|\omega_n|^2} & \\ -\left(1+\bar{\omega}_n\right)B_{n-1}u_n & D_{n-1}^{(3)}\left(\Omega_{n-1}-D_{n-1}^{(1)}\right)^{-1}\end{pmatrix} && (285)\\
& \times\begin{pmatrix} \frac{\sqrt{1-|\omega_n|^2}}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}\end{pmatrix} && (286)\\
= \;& \begin{pmatrix} \frac{1+\omega_n}{1+\bar{\omega}_n} & \\ & D_{n-1}^{(3)}D_{n-1}^{(2)}D_{n-1}^{(3)}\end{pmatrix} && (287)\\
& -\begin{pmatrix} \frac{1-|\omega_n|^2}{1+\bar{\omega}_n} & \\ -\sqrt{1-|\omega_n|^2}B_{n-1}u_n & D_{n-1}^{(3)}\left(\Omega_{n-1}-D_{n-1}^{(1)}\right)^{-1}D_{n-1}^{(3)}\end{pmatrix} && (288)\\
= \;& \begin{pmatrix} \omega_n & \\ \sqrt{1-|\omega_n|^2}B_{n-1}u_n & A_{n-1}\end{pmatrix}, && (289)
\end{aligned}
$$

which implies that A_n satisfies the updating formula, and therefore established the validity of the formula.

□

Moreover, Ω has low grade which implies a band fraction representation, we explicitly construct the band matrices in the following lemma and then make use of them to construct band matrices for the TIB pair (A, B).

Lemma 27. *Suppose we have unit length vectors $u_0, u_1, \ldots, u_n \in \mathbb{C}^{p\times 1}, p < n$,*

then there exist two band matrices M_u and N_u such that

$$N_u = M_u \Omega, \tag{290}$$

where Ω has the form

$$\Omega = tril(\begin{pmatrix} u_n^* \\ u_{n-1}^* \\ \vdots \\ u_0^* \end{pmatrix} \begin{pmatrix} u_n & u_{n-1} & \cdots & u_0 \end{pmatrix}), \tag{291}$$

Proof. For $i = 0, 1, \ldots, n-p$, we know that

$$\begin{pmatrix} u_{p+i} & u_{p-1+i} & \cdots & u_{i+1} & u_i \end{pmatrix} \in \mathbb{C}^{p \times (p+1)} \tag{292}$$

is singular, so we can find $\begin{pmatrix} v_{p,i} \\ v_{p-1,i} \\ \vdots \\ v_{1,i} \\ v_{0,i} \end{pmatrix} \in \mathbb{C}^{(p+1)\times 1}$ such that

$$\begin{pmatrix} u_{p+i} & u_{p-1+i} & \cdots & u_{i+1} & u_i \end{pmatrix} \begin{pmatrix} v_{p,i} \\ v_{p-1,i} \\ \vdots \\ v_{1,i} \\ v_{0,i} \end{pmatrix} = 0. \tag{293}$$

Let

$$M_u = \begin{pmatrix} 1 & & & & & & & & \\ 0 & 1 & & & & & & & \\ \vdots & \ddots & 1 & & & & & & \\ 0 & \cdots & 0 & 1 & & & & & \\ v^*_{p,n-p-1} & \cdots & \cdots & v^*_{1,n-p-1} & v^*_{0,n-p-1} & & & & \\ & \ddots & & & \ddots & \ddots & & & \\ & & v^*_{p,1} & & & v^*_{1,1} & v^*_{0,1} & & \\ & & & v^*_{p,0} & \cdots & \cdots & v^*_{1,0} & v^*_{0,0} \end{pmatrix} \tag{294}$$

then

$$N_u = M_u \Omega \tag{295}$$

is a lower triangular matrix with bandwidth p.

□

Remark 28. For the vectors $\begin{pmatrix} v_{p,i} \\ v_{p-1,i} \\ \vdots \\ v_{1,i} \\ v_{0,i} \end{pmatrix}$ in the proof above, it's entirely possible that the last element $v_{0,j}$ is zero, in which case M_u is singular. If all the $v_{0,0}, v_{0,1}, \ldots, v_{0,n-p-1}$ are not zero, M_u is nonsingular. Also we know the lower triangular matrix Ω has ones on its diagonal so it's non-singular, therefore Ω can be written as band ratio,

$$\Omega = M_u^{-1} N_u. \tag{296}$$

Theorem 29. *There exist band matrices M and N, such that for any TIB pair (A, B) satisfying the non-singularity condition described in remark (28), we have $A = M^{-1}N$, and $B = M^{-1}\begin{bmatrix} U \\ 0 \end{bmatrix}$ with some unitary matrix U.*

Proof. Using the notation above, if we furthur assume M_u in lemma (27) is

nonsingular, by theorem (26) we have

$$
\begin{aligned}
A &= D^{(3)}\left(D^{(2)}-\left(M_u^{-1}N_u-D^{(1)}\right)^{-1}\right)D^{(3)} && (297)\\
&= D^{(3)}\left(D^{(2)}-\left(N_u-M_uD^{(1)}\right)^{-1}M_u\right)D^{(3)} && (298)\\
&= D^{(3)}\left(N_u-M_uD^{(1)}\right)^{-1}\left(N_uD^{(2)}-M_uD^{(1)}D^{(2)}-M_u\right)D^{(3)} && (299)\\
&= \left[\left(N_u-M_uD^{(1)}\right)D^{(3)-1}\right]^{-1} && (300)\\
&\quad\times\left[\left(N_uD^{(2)}-M_uD^{(1)}D^{(2)}-M_u\right)D^{(3)}\right] && (301)
\end{aligned}
$$

and

$$
\begin{aligned}
B &= D^{(3)}\left(M_u^{-1}N_u-D^{(1)}\right)^{-1}\begin{pmatrix}u_n^*\\u_{n-1}^*\\\vdots\\u_0^*\end{pmatrix} && (302)\\
&= \left[\left(N_u-M_uD^{(1)}\right)D^{(3)-1}\right]^{-1}\left(M_u\begin{pmatrix}u_n^*\\u_{n-1}^*\\\vdots\\u_0^*\end{pmatrix}\right) && (303)\\
&= \left[\left(N_u-M_uD^{(1)}\right)D^{(3)-1}\right]^{-1}\begin{pmatrix}u_n^*\\\vdots\\u_{n-p+1}^*\\0\\\vdots\\0\end{pmatrix}. && (304)
\end{aligned}
$$

Finally let

$$
\begin{aligned}
M &= \left(N_u - M_u D^{(1)}\right) D^{(3)-1} && (305)\\
&= N_u \begin{pmatrix} \frac{1+\bar{w}_n}{\sqrt{1-|w_n|^2}} & & \\ & \ddots & \\ & & \frac{1+\bar{w}_0}{\sqrt{1-|w_0|^2}} \end{pmatrix} && (306)\\
&\quad -M_u \begin{pmatrix} \frac{\bar{w}_n}{\sqrt{1-|w_n|^2}} & & \\ & \ddots & \\ & & \frac{\bar{w}_0}{\sqrt{1-|w_0|^2}} \end{pmatrix} && (307)\\
&= (N_u + (N_u - M_u)\, diag(\bar{w}))\, diag(\frac{1}{\sqrt{1-|w|^2}}), && (308)
\end{aligned}
$$

and

$$
\begin{aligned}
N &= \left(N_u D^{(2)} - M_u D^{(1)} D^{(2)} - M_u\right) D^{(3)} && (309)\\
&= N_u \begin{pmatrix} \frac{1+w_n}{\sqrt{1-|w_n|^2}} & & \\ & \ddots & \\ & & \frac{1+w_0}{\sqrt{1-|w_0|^2}} \end{pmatrix} && (310)\\
&\quad -M_u \begin{pmatrix} \frac{1}{\sqrt{1-|w_n|^2}} & & \\ & \ddots & \\ & & \frac{1}{\sqrt{1-|w_0|^2}} \end{pmatrix} && (311)\\
&= (N_u diag(w) + (N_u - M_u))\, diag(\frac{1}{\sqrt{1-|w|^2}}). && (312)
\end{aligned}
$$

Then clearly M and N are band matrices, and we have

$$
\begin{aligned}
A &= M^{-1}N, && (313)\\
B &= M^{-1} \begin{pmatrix} u_n^* \\ \vdots \\ u_{n-p+1}^* \\ 0 \\ \vdots \\ 0 \end{pmatrix}, && (314)
\end{aligned}
$$

and $\begin{pmatrix} u_n^* \\ \vdots \\ u_{n-p+1}^* \end{pmatrix}$ is unitary.

□

Remark 30. The SISO case band fraction is the special case of the MIMO case. In the SISO case, the unit length null vectors all become scalar 1,

$$\begin{pmatrix} u_n & u_{n-1} & \cdots & u_0 \end{pmatrix} = \begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}, \tag{315}$$

then

$$\Omega = tril\left(\begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}\begin{pmatrix} 1 & 1 & \cdots & 1 \end{pmatrix}\right) \tag{316}$$

$$= \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & 1 & \\ & & 1 & 1 \end{pmatrix}^{-1}\begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix} \tag{317}$$

gives

$$M_u = \begin{pmatrix} 1 & & & \\ 1 & 1 & & \\ & \ddots & 1 & \\ & & 1 & 1 \end{pmatrix} \tag{318}$$

and

$$N_u = \begin{pmatrix} 1 & & & \\ & 1 & & \\ & & \ddots & \\ & & & 1 \end{pmatrix}. \tag{319}$$

Therefore A is a band fraction with bandwidth 1 and B is the first column of the inverse of band matrix with bandwidth 1.

The real system feedback matrix A can have complex conjugate pairs of eigenvalues, in which case, the band fraction representation just described would

be complex. Computationally, this may not be desirable, thus we give a real band fraction form for the complex conjugate poles case. Suppose that we have an input balanced pair (A, B), where A, B have real values. Let's apply the real Schur decomposition introduced in section 2.5 to A,

$$A = QTQ^*, \tag{320}$$

where T is a lower quasi-triangular matrix and Q is unitary, both T and Q only have real elements . Without loss of generality suppose that T has k 2×2 blocks on the left upper side of the diagonal,

$$T = \begin{pmatrix} T_1 & & & \\ * & \ddots & & \\ \vdots & \ddots & T_k & \\ * & \cdots & * & T_r \end{pmatrix} \tag{321}$$

where $T_1, T_2, \ldots, T_k$ are 2×2 real blocks and T_r is strictly lower triangular. For $T_k, k = 1, 2, \ldots k$, it's trivial to find its LQ decomposition,

$$T_k = L_k G_k, \tag{322}$$

where L_k are 2×2 lower triangular matrices and G_k are 2×2 unitary matrices. Denote

$$G = \begin{pmatrix} G_1 & & & \\ & \ddots & & \\ & & G_k & \\ & & & I \end{pmatrix}, \tag{323}$$

then we have

$$T = LG, \tag{324}$$

where L is strictly lower triangular and G is a unitary matrix with 2×2 unitary blocks and identity on its diagonal. The input balance condition yields

$$AA^* + BB^* = I, \tag{325}$$

which is

$$QLGQ^*QG^*L^*Q^* + BB^* = I, \tag{326}$$

and therefore

$$LL^* + (Q^*B)(Q^*B)^* = I. \tag{327}$$

Now we have a real TIB pair (L, Q^*B), according to theorem (29), there exists a real band fraction representation of the new TIB pair. All we need to do is post multiply by the unitary matrix G.

6 Model Identification and Reduction

6.1 Model Reduction Technique Review

Balanced realizations are well-known to have numerical advantages and are useful for model reduction purposes in conjunction with balance-and-truncate type procedures.

A balanced canonical form for SISO stable all-pass systems in discrete time has a positive upper triangular reachability matrix. In the multivariable case, Kronecker indices and nice selections are used to arrive at balanced overlapping canonical forms for lossless systems. For discrete-time stable all-pass systems, canonical forms can be obtained from the results in continuous-time by application of bilinear transformation. However, this destroys certain nice properties of the canonical form; e.g., truncation of state components no longer leads to reduced order systems that are balanced and in canonical form.

Let Σ be the SISO discrete time model. Two discrete-time Lyapunov equations are closely related to this system:

$$APA^* + BB^* = P, \tag{328}$$

and

$$A^*QA + C^*C = Q. \tag{329}$$

Under the assumptions that Σ is asympototically stable and minimal, it is well known that the above equations have unique symmetric positive definite solutions $P, Q \in \mathbb{R}^{n\times n}$, called the reachability and obserbility Grammians, respectively. The square roots of the eigenvalues of the product PQ are the singular values of the Hankel operator associated with Σ and are called Hankel Singular values $\sigma_i(\Sigma)$ of the system Σ:

$$\sigma_i(\Sigma) = \sqrt{\lambda_i(PQ)}. \tag{330}$$

The minimal and asymptotically stable system Σ is called balanced if P and Q are identical and diagonal with the largest Hankel singular values. The balanced system has the property that the states which are difficult to reach, i.e. require a large input energy to reach, are simultaneously difficult to observe, i.e. yield small observation energy. The states which have this property correspond to small Hankel singular values. Hence a reduced model is simply obtained by

truncating these states from the balanced system. The reduced system Σ_r defined by balanced truncation is asymptotically stable, and the error system satisfies the following $\mathcal{H}_\infty$ error bound:

$$\|\Sigma - \Sigma_r\|_{\mathcal{H}_\infty} \leq 2\left(\sigma_{k+1} + \cdots + \sigma_q\right). \tag{331}$$

Reduced-order models obtained by balanced truncation have certain guaranteed properties. However these properties are slightly different for discrete-time and continuous-time systems. For discrete-time systems, the reduced systems are not balanced in general. D.Hinrichsen and A.J.Pritchard [33] presented a direct proof of the stricter estimate for the discrete-time case under substantially weaker conditions. The authors proposed that "the reduced order model Σ_r will be constructed by projection". For continuous-time case, suppose that the positive definite matrices P and Q have the decompositions

$$P = UU^T \tag{332}$$

and

$$Q = LL^T. \tag{333}$$

Let

$$U^T L = ZSY^T \tag{334}$$

be the singular value decomposition. Define

$$W_1 = LY_1\Sigma_1^{-1/2} \tag{335}$$

and

$$V_1 = UZ_1\Sigma_1^{-1/2}, \tag{336}$$

where Z_1 and Y_1 are composed of the leading r columns of Z and Y, respectively. It is easy to check that

$$W_1^T V_1 = I_r \tag{337}$$

and hence that $V_1 W_1^T$ is an oblique projector. We obtain a reduced model Σ_r

of order r by projection as follows:

$$A_r = W_1^T A V_1, \tag{338}$$

$$B_r = W_1^T B, \tag{339}$$

$$C_r = C V_1. \tag{340}$$

Noting that $PW_1 = V_1\Sigma_1$ and $QV_1 = W_1\Sigma_1$ gives

$$W_1^T(AP + PA^T + BB^T)W_1 = A_r\Sigma_1 + \Sigma_1 A_r^T + B_r B_r^T, \tag{341}$$

and

$$V_1^T(A^TQ + QA + C^TC)V_1 = A_r^T\Sigma_1 + \Sigma_1 A_r + C_r^T C_r. \tag{342}$$

Thus, the reduced model is balanced and asympototically stable(due to the *Lyapunov inertia theorem*) for any $k \leq q$. However, in contrast to the continuous-time case, the reduced order models are not necessarily balanced.

Theorem 31. *From [5], if* $\deg \hat{H}(z) = M$, *then* $\hat{H}(z)$ *is a stationary point of the functional* $\|H(z) - \hat{H}(z)\|_2$ *if and only if*

$$H(z) - \hat{H}(z) = z[V(z)]^2 Q(z), \ \textit{for some } Q(z) \in \mathcal{H}_2, \tag{343}$$

where $V(z)$ *is the all-pass function whose poles coincide with those of* $\hat{H}(z)$.

This result was first obtained by Walsh, and hence is known in some circles as Walsh's theorem [5]. The essence of this theorem is best intepreted as an interpolation condition. If $z_1, \ldots, z_M$ are the poles of $\hat{H}(z)$, and if the poles are distinct, this interpolation constraint reads as

$$H(\frac{1}{z_k}) = \hat{H}(\frac{1}{z_k}), \tag{344}$$

and

$$\frac{\partial H(z)}{\partial z} \mid_{z=\frac{1}{z_k}} = \frac{\partial \hat{H}(z)}{\partial z} \mid_{z=\frac{1}{z_k}}, \tag{345}$$

for $k = 1, 2, \ldots, M$. The result above is revisited by Regalia [5].

A.Bunse-Gerstner et al. [34] derived the first order necessary H_2-optimality conditions (Theorem 32) for asymptotically stable MIMO systems. The mirror images of the eigenvalues of the state matrix $\hat{A}$ are crucial quantities herein.

Theorem 32. *From [34], given the large order system with transfer function $H(s)$. Let $\hat{H}(s)$ be the transfer function of the reduced order system given in an eigenvector basis*

$$\begin{aligned} \hat{A} &= diag(\hat{\lambda}_1, \ldots, \hat{\lambda}_n), && (346) \\ \hat{B} &= \left(\hat{b}_1^*, \ldots, \hat{b}_n^*\right), && (347) \\ \hat{C} &= (\hat{c}_1, \ldots, \hat{c}_n). && (348) \end{aligned}$$

If $\hat{H}(s)$ solves the H_2-optimal problem, then the following conditions are satisfied

$$\begin{aligned} \hat{c}_k^* H\left(\frac{1}{\hat{\lambda}_k^*}\right) &= \hat{c}_k^* \hat{H}\left(\frac{1}{\hat{\lambda}_k^*}\right), && (349) \\ H\left(\frac{1}{\hat{\lambda}_k^*}\right)\hat{b}_k^* &= \hat{H}\left(\frac{1}{\hat{\lambda}_k^*}\right)\hat{b}_k^*, && (350) \\ \hat{c}_k^* H'\left(\frac{1}{\hat{\lambda}_k^*}\right)\hat{b}_k^* &= \hat{c}_k^* \hat{H}\left(\frac{1}{\hat{\lambda}_k^*}\right)\hat{b}_k^*. && (351) \end{aligned}$$

In Antoine Vandendorpe's thesis [26] a generalization of existing interpolation techniques for MIMO systems is developed. Instead of imposing interpolation conditions of the type

$$T(\lambda_i) = \hat{T}(\lambda_i), \tag{352}$$

more general tangential interpolation conditions can be imposed between the original and the reduced order systems:

$$\begin{aligned} x_i T(\lambda_i) &= x_i \hat{T}(\lambda_i), && (353) \\ T(\lambda_{i+k}) y_i &= \hat{T}(\lambda_{i+k}) y_i. && (354) \end{aligned}$$

Such interpolation conditions appear naturally for MIMO systems when projecting via Sylvester equations**.**

In A.Bunse-Gerstner's paper [34] , they introduced a special case of much more general results of A.Vandendorpe.

Lemma 33. *From [34], let $V_n \in \mathbb{C}^{N\times n}$ and $W_n \in \mathbb{C}^{N\times n}$ be matrices of full*

rank n such that

$$W_n^* V_n = I_n. \tag{355}$$

Let $\sigma_k \in \mathbb{C}, l_k \in \mathbb{C}^{1\times p}$ *and* $r_k \in \mathbb{C}^{m\times 1}$ *for* $k = 1, \ldots, n$ *be given sets of interpolation points and left and right tangential directions, respectively. Assume that the points* σ_k *are chosen such that all matrices* $A - \sigma_k I_N$ *are invertible. If for all* $k \in 1, \ldots, n$,

$$\begin{aligned} \left(\sigma_k I_N - A\right)^{-1} B r_k &\in \mathit{columnspace}\left(V_n\right), & (356)\\ \left(\sigma_k^* I_N - A^*\right) C^* l_k^* &\in \mathit{columnspace}\left(W_n\right), & (357) \end{aligned}$$

then the reduced order system $\hat{\Sigma} = \left(\hat{A}, \hat{B}, \hat{C}\right) = (W_n^* A V_n, W_n^* B, C V_n)$ *has a transfer function which satisfies the following tangential interpolation conditions:*

$$\begin{aligned} H(\sigma_k) r_k &= \hat{H}(\sigma_k) r_k, & (358)\\ l_k H(\sigma_k) &= l_k \hat{H}(\sigma_k), & (359)\\ l_k H'(\sigma_k) r_k &= l_k \hat{H}'(\sigma_k) r_k. & (360) \end{aligned}$$

From Lemma 32 and Lemma 33, A-Bunse-Gerstner introduced MIRIAm(MIMO Iterative Rational Interpolation Algorithm), a proof of the convergence is not provided.

Gugencin [43] proposed a two-sided projection combining features of the singular value decomposition (SVD)-based and the Krylov-based model reduction technique. While the SVD-side of the projection depends on the observability Grammian, the Krylov-side is obtained via iterative rational Krylov steps.

Beattie and Gugercin [41] presented a trust-region approach for optimal H_2 model reduction of MIMO linear dynamical systems. They generated a sequence of reduced order models producing monotone improving H_2 error norms and is globally convergent to a reduced order model guaranteed to satisfy first-order optimality conditions with respect to H_2 error criteria without solving any Lyapunov equations. The method also appeared to be the first descent approach that uses Hessian information.

In [44], the problem of medium-scale MIMO linear time invariant (LTI) systems is studied. The author also presented a MATLAB-based toolbox for approximation of medium and large-scale LTI dynamical models, called MORE

(MOdel REduction), which implements a collection of very recent advanced algorithms for LTI dynamical model reduction purpose.

The celebrated theorem of Adamjov, Arov, and Krein (AAK), provides a construction of the optimal approximations to a Hankel operator bounded in Hankel norm. [36] is a good reference instructing how to implement AAK algorithm. [24] first found a parametrization for inner functions, then they tackled the reduction problem by using a gradient algorithm through the manifold as a whole, using the coordinate maps to describe the manifold locally and changing from one coordinate map to another when required. Our reduction algorithm is based on their parametrization.

6.2 Fast Partial Block Hankel SVD

We introduce a hybrid model reduction algorithm based on the TIB representation: first we obtain the bases of the lossless function factor from the H^∞ approximation, then the remaining unstable factor is obtained by H^2 approximation. To find the H^∞ approximation, we need to compute the partial SVD of a block Hankel matrix multiplied by its transpose, which can be accelerated by the FFT as follows.

Given a Hankel matrix

$$H = \begin{pmatrix} h_1 & h_2 & \cdots & h_n \\ h_2 & & & h_{n+1} \\ \vdots & & & \vdots \\ h_n & h_{n+1} & \cdots & h_{2n-1} \end{pmatrix}, \tag{361}$$

we know that the fast multiplication of H by vector $x = (x_1, x_2, \ldots, x_n)^T$ can be achieved using fast Fourier transformation in $O(n \log n)$ operations. Suppose $y = Hx$, actually

$$\hat{y} = IFFT\left(FFT\left(\hat{h}\right) * FFT\left(\hat{x}\right)\right), \tag{362}$$

where

$$\hat{h} = (h_n, h_{n+1}, \ldots, h_{2n-1}, h_1, \ldots, h_{n-1})^T, \tag{363}$$

$$\hat{x} = (x_n, x_{n-1}, \ldots, x_1, 0, \ldots, 0)^T, \tag{364}$$

and

$$\hat{y} = (y_1, y_2, \ldots, y_n, \ldots)^T . \tag{365}$$

For a block Hankel matrix multiplication, suppose that we have block Hankel matrix,

$$H = \begin{pmatrix} H_1 & H_2 & \cdots & H_n \\ H_2 & & & \vdots \\ \vdots & & & H_{n+1} \\ H_n & H_{n+1} & \cdots & H_{2n-1} \end{pmatrix}, \tag{366}$$

where $H_i \in \mathbb{R}^{p\times q}$ for $i = 1, 2, \ldots, 2n-1$. For the following, we adopt the index notation of Matlab. Note that $H_{i:p:end,j:q:end}$ are Hankel matrices for every $i = 1, 2, \ldots, p$ and $j = 1, 2, \ldots, q$. For vector $x \in \mathbb{R}^{nq}$, and $y = Hx, y \in \mathbb{R}^{np}$,

$$y_{i:p:end} = \sum_{j=1}^{q} H_{i:p:end,j:q:end} x_{j:q:end} \tag{367}$$

which is the sum of q Hankel multiplication. Notice that in the block Hankel case, H does not have to be symmetric. However, H^*H is Hermitian and semi-positive-definite, so its singular values are the same as its eigenvalues. Utilizing the fast multiplication of H^*H, we can use the Lanczos algorithm to compute the partial-decomposition with the largest eigenvalues in terms of magnitude. Suppose that we have a partial SVD of H of rank γ,

$$H \approx \widetilde{H} = \widetilde{U}\widetilde{S}\widetilde{V}^* \tag{368}$$

where $\widetilde{S} \in \mathbb{R}_+^{\gamma\times\gamma}$. Then $H^*H \approx \widetilde{V}\widetilde{S}^2\widetilde{V}^*$ and $HH^* \approx \widetilde{U}\widetilde{S}^2\widetilde{U}^*$, by applying Lanczos algorithm to H^*H and HH^*, we will the SVD approximation of H.

6.3 Hybrid Model Reduction with TIB

Model reduction of linear time-invariant systems is an active research area, and traditional methods include balanced truncation and moment matching; the AAK algorithm will find the optimal solution in the H^∞ sense [8]. More recent model reduction algorithms include oblique projection combining the aspects of the SVD and Krylov based reduction methods [10, 11], and gradient algorithms based on Schur analysis [7, 12]. The algorithm we propose is based on the latter, however, in place of optimization over local manifolds with changing coordinate maps, we find the optimal inner part of the transfer function in the H^∞ sense,

and then compute the rest part with nice geometry. Advantages of the algorithm include that it is fast, it is guaranteed to converge, and it is stable. For transfer functions in VMOA (Space of analytic functions of Vanishing Mean Oscillation) [62], it is an accurate approximation.

Since transfer functions in VMOA are approximated in Hankel norm by their truncations, without loss of generality we assume the transfer function is of finite (but possibly very long) length k. Transfer functions in VMOA have compact Hankel operators. The Hankel operator is approximated in Hankel norm by the finite rank partial SVDs. For the rank γ partial SVD of $H \approx \widetilde{H} = \widetilde{U}\widetilde{S}\widetilde{V}^*$, there exists a state space realization pair (A, B) such that,

$$\tilde{V}^* \approx \begin{pmatrix} B & AB & \cdots & A^{k-1}B \end{pmatrix} \tag{369}$$

and for k large,

$$\tilde{V}^*_{1:end,q+1:end}\tilde{V}_{1:end,1:end-p} \tag{370}$$

$$= \begin{pmatrix} AB & A^2B & \cdots & A^{k-1}B \end{pmatrix} \begin{pmatrix} B^* \\ B^*A^* \\ \vdots \\ B^*A^{(k-2)*} \end{pmatrix} \tag{371}$$

$$= A\left(BB^* + ABB^*A + \cdots A^{k-2}BB^*A^{(k-2)*}\right) \tag{372}$$

$$\approx A \tag{373}$$

and $B = \tilde{V}_{1:end,1:q}$. As the impulse response is preserved under a linear transformation of the representation,

$$A \rightarrow TAT^{-1} \tag{374}$$

$$B \rightarrow TB \tag{375}$$

$$C \rightarrow CT^{-1}, \tag{376}$$

we may consider the Schur triangularization of A,

$$A = QA_1Q^T, \tag{377}$$

where A_1 is lower triangular, and observe that $\left(A_1, Q^TB\right)$ is an equivalent realization pair. However,$\left(A_1, Q^TB\right)$ is not in general a TIB pair, so we will find

a TIB approximation by tangential Schur updating formula. The reduced poles can be taken as the diagonal elements of A_1, which we denote as ω, and the null vectors u can be recursively computed as

$$u_i = B^*_{1,1:end}/\|B_{1,1:end}\|, \tag{378}$$

$$newB = B_{2:end,1:end} - \left(1 + \frac{1}{\bar{w}_i}\right) B_{2:end,1;end} u_i u_i^*. \tag{379}$$

Finally we can use ω and u to reconstruct a strict TIB pair, which determine the inner function and provide the orthonormal reduced basis for the matrix-valued transfer function. In this reduced basis H^2 approximation of C is a well conditioned least square problem. We compute the TIB pair $\left(\widetilde{A}, \widetilde{B}\right)$ from ω and u and the Krylov matrix:

$$K = \begin{pmatrix} B & AB & \cdots & A^{k-1}B \end{pmatrix}, \tag{380}$$

and find

$$C = \begin{pmatrix} H_1 & H_2 & \cdots & H_k \end{pmatrix} K^*. \tag{381}$$

To summarize the reduction algorithm, we have the following 4 steps:

1. Partial SVD approximation of Hankel matrix $H \approx \widetilde{H} = \widetilde{U}\widetilde{S}\widetilde{V}^*$.
2. Get almost input balanced pair from $\tilde{V}^* \approx \begin{pmatrix} B & AB & \cdots & A^{k-1}B \end{pmatrix}$ and apply coordinate change to get almost TIB pair.
3. Compute poles and null vectors from the almost TIB pair by recursive tangential Schur algorithm.
4. Reconstruct strict TIB pair from estimated poles and null vectors, then conduct least square method to obtain C.

For impulse responses in VMOA, the truncation and partial SVD are guaranteed to be close in the Hankel norm. In this case, the approximation minimizes the H_2 error of the impulse response subject to minimal Hankel norm error of the inner part. Since the Hankel norm dominates the H_2 norm the process guarantees a small H_2 norm, but not necessarily minimal H_2 norm for the given γ. Numerical tests of this algorithm on synthetic data, are given in section 6.5.

It is appealing to conduct the reduction in the information space. From the information geometric point of view, in the SISO case, we know that if we choose the model parameters to be the coefficients of the logarithm of the transfer function, the Fisher information matrix is an identity matrix indicating a nice Euclidean statistical manifold. Denoting $\log f(z) = a_0 + a_1 z + a_2 z^2 + \cdots$, since f and $\log f$ have the same singularity, we can apply our reduction algorithm on the alternative Hankel matrix

$$A_{ij} = a_{i+j-1}, i, j = 0, 1, \ldots \tag{382}$$

and then recover the impulse responses from $\{a_i\}$. Mullhaupt and Choi [61] proved in the SISO case, the information of prediction only comes from the unstable part of the transfer function, thus we have reason to believe that the Douglas-Shapiro-Shields factorization plays a similar role in matricial information geometry.

6.4 Model Idenfication with TIB

We consider adaptive identification of the impulse response of an innovation filter,

$$\begin{aligned} z(t+1) &= Az(t) + Bx(t) && (383) \\ y(t) &= Cz(t) + x(t). && (384) \end{aligned}$$

Innovation models use the prediction fit errors as the stochastic input into the state space evolution. We assume that the system is minimal and stable, in addition we will choose the TIB representation for the state space system. The innovations are independent and identically distributed with zero mean and variance σ^2. From equation (383) we have

$$z(t) = \begin{pmatrix} B & AB & A^2B & \cdots \end{pmatrix} \begin{pmatrix} x(t-1) \\ x(t-2) \\ x(t-3) \\ \vdots \end{pmatrix}, \tag{385}$$

all the historical information can be encoded in the state $z(t)$. The auto-

covariance of $z(t)$ is

$$
\begin{aligned}
& E\left[z(t)\, z(t)^*\right] && (386)\\
= \; & E\left[\begin{pmatrix} B & AB & \cdots \end{pmatrix}\begin{pmatrix} x(t-1) \\ x(t-2) \\ \vdots \end{pmatrix}\begin{pmatrix} x(t-1) & x(t-2) & \cdots \end{pmatrix}\begin{pmatrix} B^* \\ B^*A^* \\ \vdots \end{pmatrix}\right] && (387)\\
= \; & \begin{pmatrix} B & AB & \cdots \end{pmatrix} E\left[\begin{pmatrix} x(t-1) \\ x(t-2) \\ \vdots \end{pmatrix}\begin{pmatrix} x(t-1) & x(t-2) & \cdots \end{pmatrix}\right]\begin{pmatrix} B^* \\ B^*A^* \\ \vdots \end{pmatrix} && (388)\\
= \; & \begin{pmatrix} B & AB & \cdots \end{pmatrix}\sigma^2 I\begin{pmatrix} B^* \\ B^*A^* \\ \vdots \end{pmatrix} && (389)\\
= \; & \sigma^2 I, && (390)
\end{aligned}
$$

when (A, B) is input balanced. By multiplying equation (384) by $z(t)^*$ and taking expecation on both sides,

$$E\left[y(t)\, z(t)^*\right] = E\left[Cz(t)\, z(t)^*\right] + E\left[x(t)\, z(t)^*\right], \tag{391}$$

since $E\left[x(t)\, z(t)^*\right] = 0$,

$$E\left[y(t)\, z(t)^*\right] = CE\left[z(t)\, z(t)^*\right], \tag{392}$$

which gives us the estimation of C,

$$C = \frac{1}{\sigma^2} E\left[y(t)\, z(t)^*\right] \tag{393}$$

The TIB pair (A, B) gives an orthogonal basis for the linear system, in practice, we assume that it changes infrequently and put all the adaptiveness into C. Adaptive filtering is gaining favor in numerous applications to help cope with time-variations of system parameters, and to compensate for the lack of a prior knowledge of the statistical properties of the input data. Over the last several years, a wide range of algorithms has been developed. These fall into four main groups [2, 5, 13, 40]: recursive least squares (RLS) algorithms and the corresponding fast versions; QR- and Inverse QR-least squares algorithms; least-squares lattice (LSL) and QR decomposition-based least squares lattice

(QRD-LSL) algorithms; and gradient-based algorithms such as the least-mean square (LMS) algorithm.

At time t, we have

$$C_t = \frac{1}{\sigma^2} E\left[y\left(t\right) z\left(t\right)^*\right] \tag{394}$$

which can be used for prediction at time $t+1$,

$$E\left[y\left(t+1\right)\right] = C_t z\left(t\right). \tag{395}$$

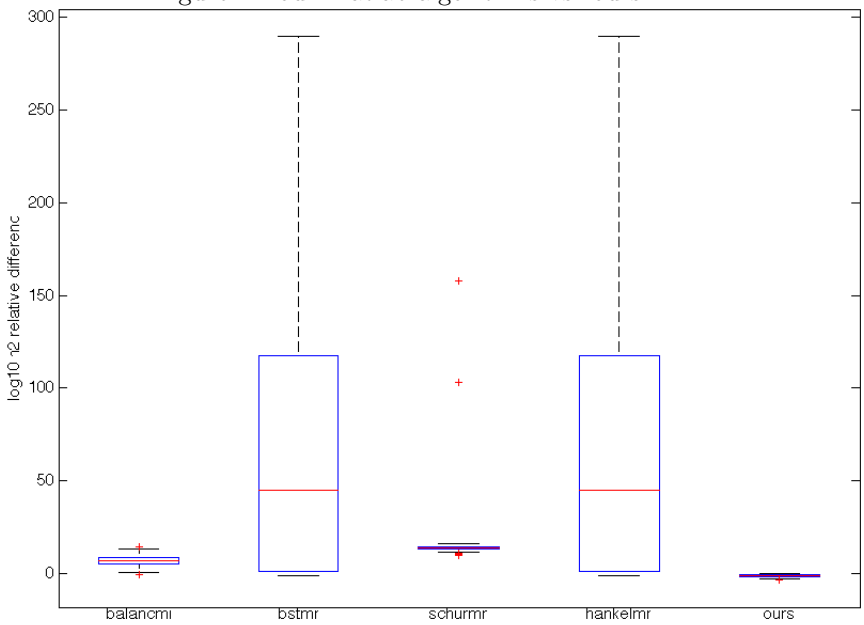
How to choose a "good" basis for the system becomes very import. In SISO case, practically we can choose some non-informative poles, for example Chebyshev poles that have the property to mitigate the Runge's phenomenon, to construct a basis for learning the underlying impulse response, then apply a model reduction algorithm to find a set of reduced poles that can span the space the impulse reponse lies onto. The reduction also serves as noise reduction duty, after getting the reduced poles, we can re-learn the impulse reponse from the reduced TIB pair. In MIMO case, a selection of non-informative null vectors must be additionally made. One example of a method for selecting such null vectors can be found in [63]. With a prior knowledge of the basis, or the poles and null vectors, we are able to identify the underlying system.

6.5 Numerical Examples

We present four numerical examples here: the first one is reducing Gilbert realization of all-pass system, our algorithm verses Matlab; the second one is finding known poles, and finding realization that approximates known impulse response; the third one takes the challenge of $1/f$ noise model reduction; the fourth one tries to approximate 2×2 maxflat impulse response and attempts to conduct model identification with reduced basis assuming we know the "real" system. We have matlab code for the algorithm proposed which takes the impulse response and number of dimension of the reduced system, and returns reduced poles, null vectors and C (C in the state space representation). With the poles and null vectors we are able to reconstruct TIB pair (A, B), and therefore the entile state space representation.

Example 34. We randomly generated 200 poles and null vectors of size 2, which will give us 200 all-pass transfer functions respectively. We choose the Gilbert realizations of these all-pass transfer functions which have diagonal A, four built-in Matlab algorithms in the Robust Control Toolbox are used to

Figure 2: four Matlab algorithms vs. ours



reduce the system from the Gilbert realizations. We also compute the length 700 impulse responses from those realizations and reduce the system by our hybrid algorithm. For all five algorithms, we reduce the systems from dimension 20 to dimension 10. Figure 2 is a boxplot of the log10 relative H_2 difference of the five reduction algorithms. Figure 3 is the comparison between the winner of Matlab algorithms and ours, by examinating the magnitude we see Matlab algorithms completely fail to solve the problem while our hybrid algorithm provides reliable solutions.

Example 35. Let's consider a 8×8 matrix-valued transfer function

$$f(z) = \begin{pmatrix} \frac{1}{1-\lambda_1 z} & \frac{1}{1-\lambda_2 z} & \cdots & \frac{1}{1-\lambda_8 z} \\ \frac{1}{1-\lambda_9 z} & \frac{1}{1-\lambda_{10} z} & \cdots & \frac{1}{1-\lambda_{16} z} \\ \vdots & \vdots & & \vdots \\ \frac{1}{1-\lambda_{57} z} & \frac{1}{1-\lambda_{58} z} & \cdots & \frac{1}{1-\lambda_{64} z} \end{pmatrix}, \tag{396}$$

with random generated poles $\lambda_1, \lambda_2, \ldots, \lambda_{64}$ inside the unit circle, the impulse

Figure 3: Matlab winner vs. ours

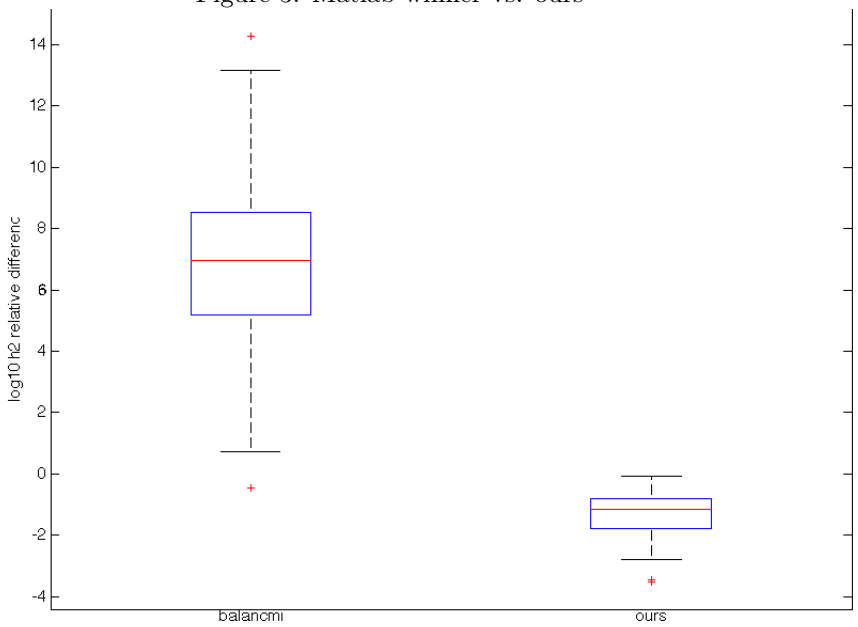
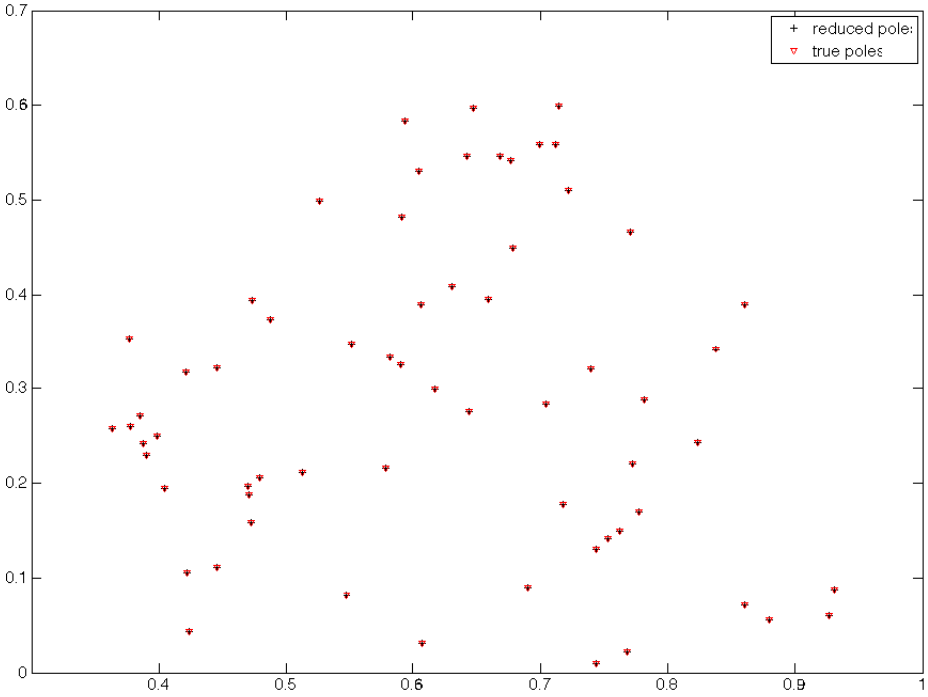


Figure 4: reduce to 64 poles



response is

$$f_i = \begin{pmatrix} \lambda_1^i & \lambda_2^i & \cdots & \lambda_8^i \\ \lambda_9^i & \lambda_{10}^i & \cdots & \lambda_{16}^i \\ \vdots & \vdots & & \vdots \\ \lambda_{57} & \lambda_{58} & \cdots & \lambda_{64} \end{pmatrix}, i = 0, 1, 2, \ldots \tag{397}$$

We pick a length 1000 trunctation of the infinitely long impulse response and apply the model reduction algorithm we proposed. If we set the reduced dimension as 64, which is the exact McMillan degree of the transfer function, we find all the poles: We also reduce the systems to McMillan degree $20, 40, 80, 100$. The poles plots are as follows:

Table 1 is a summary of relative H^2 differences for different choices of the number of reduced poles.

Figure 5: reduce to 20 poles

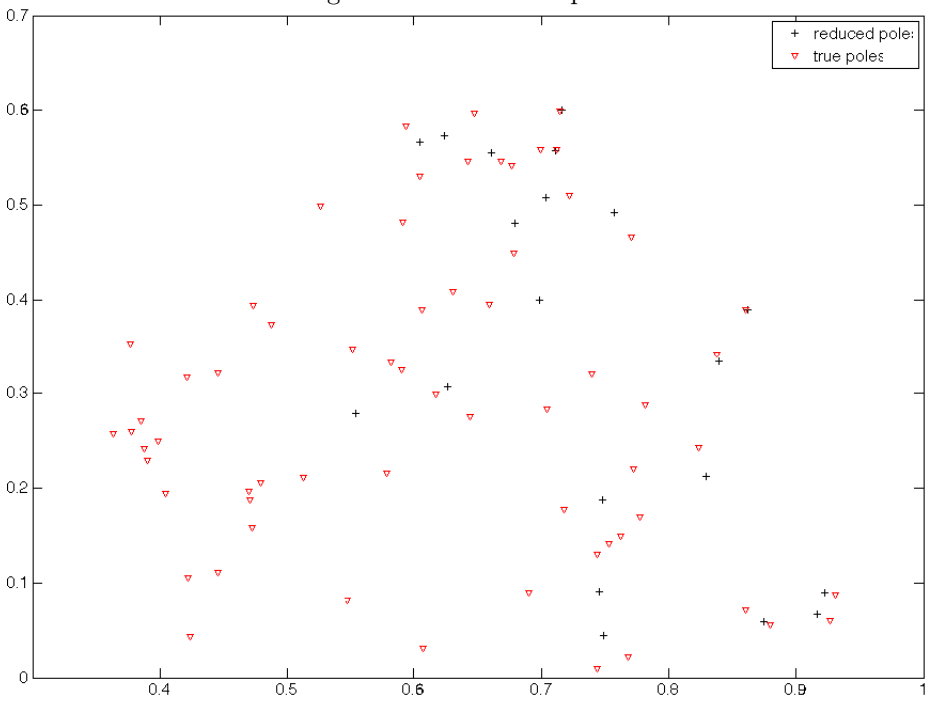


Table 1: recover poles

# of poles	real error	complex error
64	3.499833934943571e-13	3.499833934943571e-13
20	0.042523944732722	0.042523944732722
40	0.005145109028320	0.005145109028320
80	6.298718299991078e-07	6.298718299991078e-07
100	3.219230742073192e-06	3.219230742073192e-06

Figure 6: reduce to 40 poles

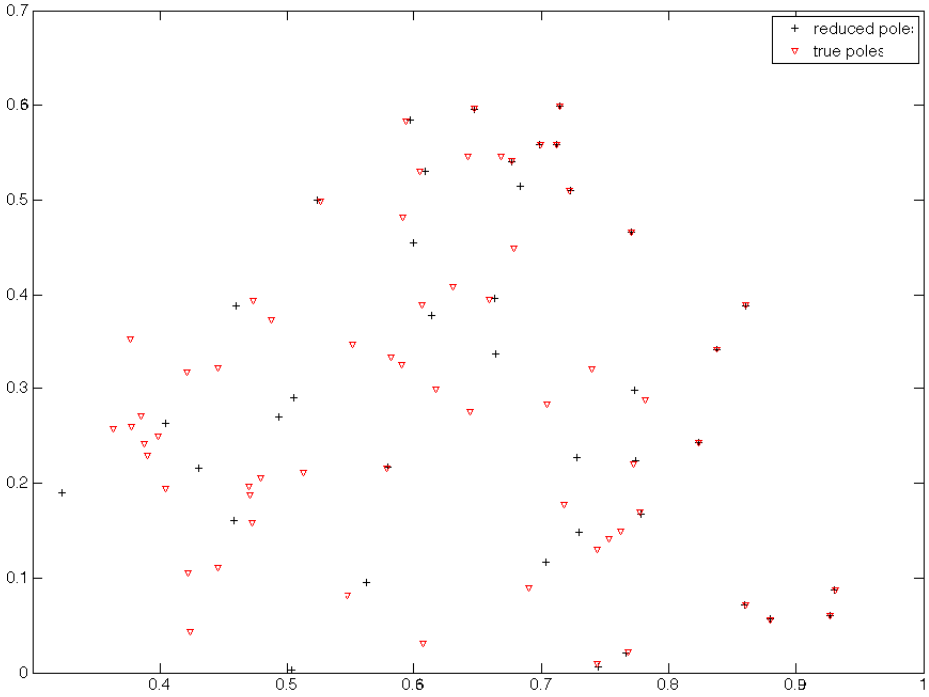


Figure 7: reduce to 80 poles

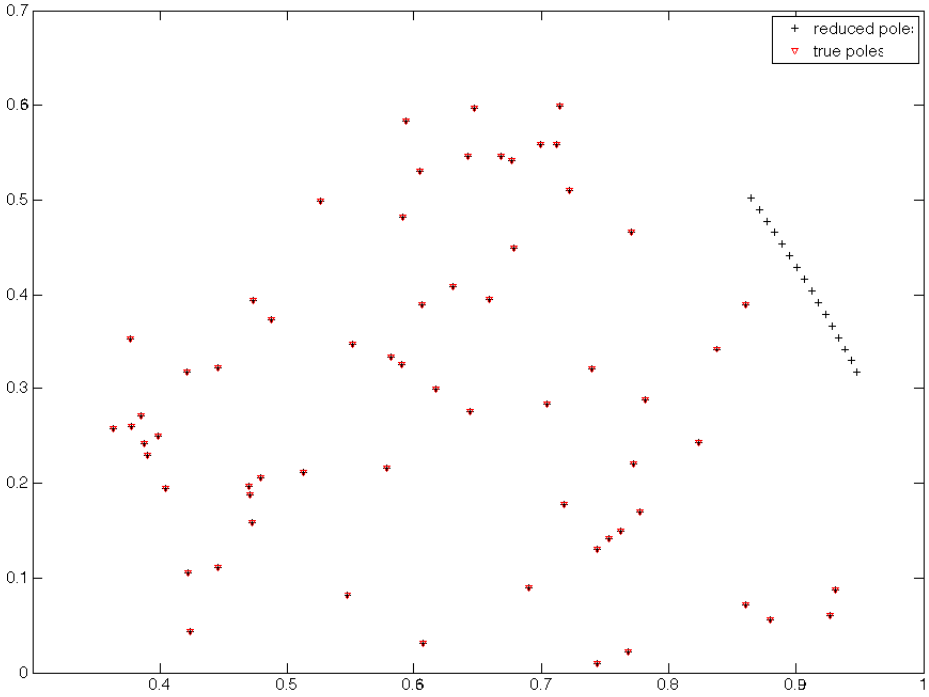


Figure 8: reduce to 100 poles

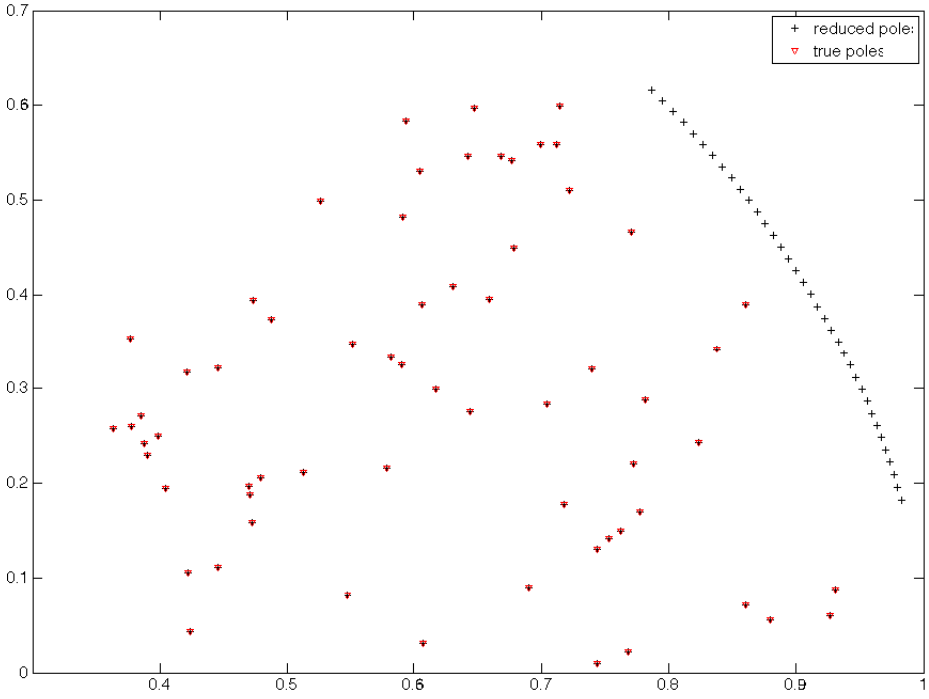


Table 2: 1/f noise approximation with different number of poles

# of poles	real error	complex error
5	0.337585415016762	0.337766019039639
10	0.125066724338163	0.125124943081321
20	0.075328146807122	0.075332050790121
30	0.058173803969318	0.058225734720307
40	0.039122621363868	0.039097241428703
50	0.027364072484523	0.027428305461233

Example 36. ($1/f$ noise) We choose the true impulse response to be

$$f(z) = \sum_{i=0}^{\infty} \begin{pmatrix} i^{-0.5} & i^{-1} \\ i^{-1.5} & i^{-2} \end{pmatrix} z^i, \tag{398}$$

all of the four entries are pink noises, which decay slowly. We pick a length 1000 trunctation of the infinitely long impulse response and apply the model reduction algorithm we proposed. We reduce the system to various dimensions, i.e. $5, 10, 20, 30, 40, 50$. The results are summarized in the Table 2, "real error" means the relative H_2 difference between true system and reduced system from real form of reduction algorithm while "complex error" means the relative H_2 difference between true system and reduced system from complex form of reduction algorithm . Here is a plot of the first 100 length of the impulse responses of 20 poles, the red plus sign is from our reduced model (Figure 9). Since the relative errors are small, other reductions look similar.

Example 37. Given a 2×2 maxflat impulse response, we compute the approximation of McMillan degree 7 obtained by our algorithm. Now 50000 sets of synthetic realizations of the impluse response are generated, namely, we take convolution of the impulse response and Gaussian distributed innovations. The data length is set as 10000, we then learn the system with the prescribed poles and null vectors reduced from true impluse reponse. Figure 10 shows the result, the blue line is the true impulse response, the green line is the reduction approximation, and the red one is the learned impulse response from the synthetic data.

Figure 9: $1/f$ noise approximation first 100 impulse response

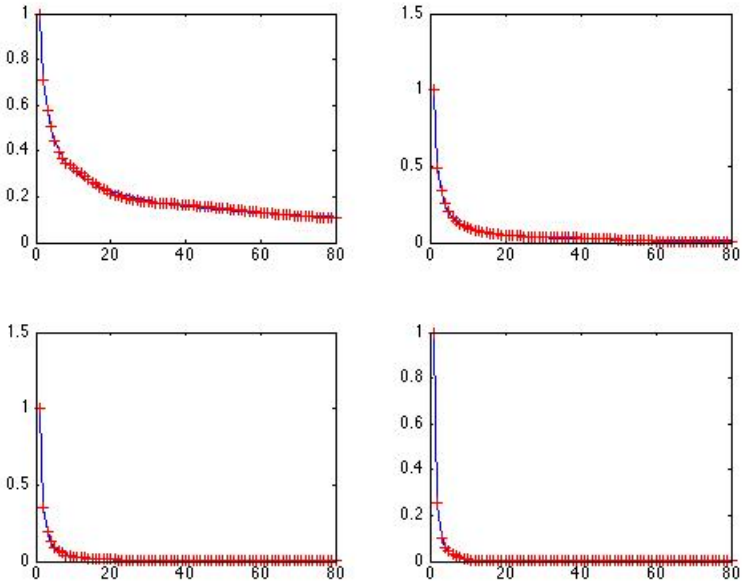
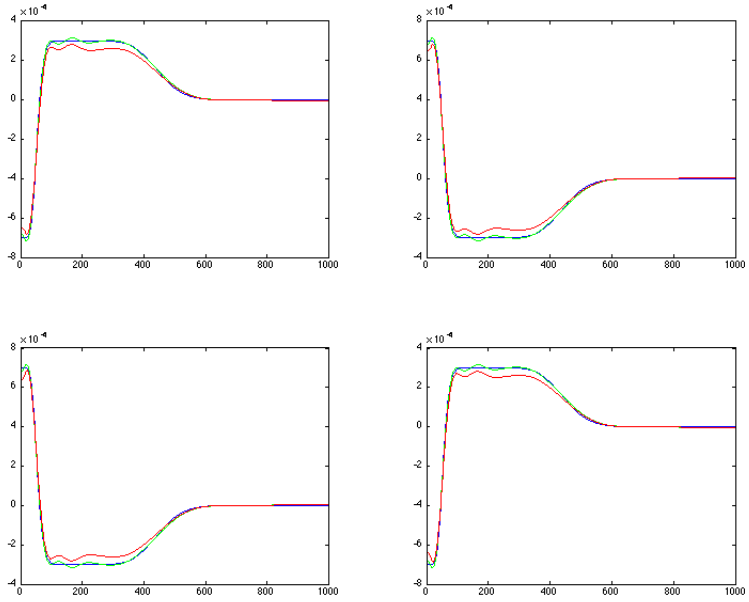


Figure 10: System identification with prescribed basis



A Appendix

A.1 An Extension of Schur-Horn Theorem

The original Schur-Horn theorem discusses the relationship between majorization and doubly stochastic matrix. Before we introduce our extension to it, we will provide the definition of majorization, doubly stochastic, biunitary and some simple properties of them.

For two vectors $a, b \in \mathbb{R}^d$ written in non-decreasing order, we say a *majorizes* b written as $a \succ b$, if and only if

$$\sum_{i=1}^{k} a_i \geq \sum_{i=1}^{k} b_i, i = 1, \ldots, d-1, \tag{399}$$

$$\sum_{i=1}^{d} a_i = \sum_{i=1}^{d} b_i. \tag{400}$$

A square matrix is B is said to be *doubly stochastic* if its matrix elements satisfy

$$B_{ij} \geq 0, \tag{401}$$

$$\sum_{i} B_{ij} = 1, \tag{402}$$

$$\sum_{j} B_{ij} = 1. \tag{403}$$

Also a square matrix B is *biunitary* if the elements of B are modulus of elements of a unitary matrix U,

$$B_{ij|} = |U_{ij}|^2. \tag{404}$$

Some mathematicians use *bistochastic* rather than doubly stochastic, or unitochastic instead of biunitary. All biunitary matrices are doubly stochastic. For 2×2 case, doubly stochastic matrices are biunitary, but not for higher dimensional cases. For example the matrix

$$A = \frac{1}{2} \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix} \tag{405}$$

is doubly stochastic. A is biunitary if and only if there exists a unitary matrix

U, such that

$$U = \frac{1}{\sqrt{2}} \begin{pmatrix} e^{ia} & e^{ib} & 0 \\ 0 & e^{ic} & e^{id} \\ e^{if} & 0 & e^{ig} \end{pmatrix}. \tag{406}$$

However,

$$\begin{aligned} UU^* &= \frac{1}{2} \begin{pmatrix} e^{ia} & e^{ib} & 0 \\ 0 & e^{ic} & e^{id} \\ e^{if} & 0 & e^{ig} \end{pmatrix} \begin{pmatrix} e^{-ia} & 0 & e^{-if} \\ e^{-ib} & e^{-ic} & 0 \\ 0 & e^{-id} & e^{-ig} \end{pmatrix} & (407) \\ &= \frac{1}{2} \begin{pmatrix} 2 & e^{i(b-c)} & e^{i(a-f)} \\ e^{i(c-b)} & 2 & e^{i(d-g)} \\ e^{i(f-a)} & e^{i(g-d)} & 2 \end{pmatrix} & (408) \end{aligned}$$

can not be identity matrix.

The set of all doubly stochastic $N \times N$ matrices is a convex set known as *Birkhoff's polytope*. A principal fact related to the *Birkhoff-von Neumann Theorem*, it states the extreme points of the Birkhoff's polytope are exactly the set of permuation matrices. The set of biunitary matrices is a subset of Birkhoff's polytope. The product of two doubly stochastic matrices is doubly stochastic, for biunitary matrices, it is not generally true that the product of two biunitary matrices is biunitary. Actually, the product of two 2×2 or 3×3 biunitary matrices is biunitary, for 4×4 or higher dimensional case, this is not necessarily true. However, if we increase the number of the factors in the product, the product appears to converge to a biunitary matrix. We study a special class of biunitary matrices which is generated by consecutive subblock product of *elementary biunitary* matrices.

Let

$$\begin{aligned} c_i &= \cos\theta_i, & (409) \\ s_i &= \sin\theta_i, & (410) \end{aligned}$$

consider the consecutive subblock product

$$A_k = \prod_{i=k}^{n} \begin{pmatrix} I_{i-k} & & & \\ & s_i & -c_i & \\ & c_i & s_i & \\ & & & I_{n-i} \end{pmatrix}, \tag{411}$$

it is a $(n-k+2)\times(n-k+2)$ upper Hessenberg unitary matrix with subdiagonal $\{c_k,\ldots,c_n\}$. Correspondingly, let

$$B_k=\prod_{i=k}^{n}\begin{pmatrix} I_{i-k} & & & \\ & s_i^2 & c_i^2 & \\ & c_i^2 & s_i^2 & \\ & & & I_{n-i}\end{pmatrix}, \tag{412}$$

which is the product of a sequence of elementary biunitary matrices generated by the factors of A_k. By induction, we can prove the value of the elements of B_k equals the square modulus of elemetns of A_k.

Lemma 38. *A_k and B_k are defined as above, then*

$$(B_k)_{ij}=|\,(A_k)_{ij}\,|^2, \tag{413}$$

immediately B_k is a biunitary upper Hessenberg matrix.

Now we present the orginal Schur-Horn theorem and our extension.

Theorem 39. *(Schur-Horn) Let $d=\{d_i\}_{i=1}^n$ and $\lambda=\{\lambda_i\}_{i=1}^n$ be real vectors in non-increasing order, then there exists a Hermitian matrix with diagonal elements d and eigenvalues λ if and only if $\lambda\succ d$.*

The equivalent form of Schur-Horn theorem is characterized by biunitary matrix. Since Hermitian matrix A has decomposition

$$A=U\Lambda U^*, \tag{414}$$

where A is with diagonal values d, U is a unitary matrix, and Λ is a diagonal matrix with λ on the diagonal. Then

$$d_i=\sum_{k=1}^{n}|u_{ik}|^2\lambda_k, i=1,\ldots,n. \tag{415}$$

Now denote $B=\left\{|u_{ij}|^2\right\}_{i.j=1}^n$, the relationship between d and λ above becomes

$$d=B\lambda. \tag{416}$$

We claim that the biunitary matrix can be made stronger to a permuated biunitary upper Hessenberg matrix.

If $\lambda \succ d$, then we can find a permuation matrix P and a biunitary upper Hessenberg matrix B, such that

$$d = BP\lambda. \tag{417}$$

We present an algorithm to construct such permuation matrix and biunitary upper Hessenberg matrix. Before the constructive proof, we first introduce *stochastic complementation* to give a heuristic explanation.

Let P be an $N \times N$ stochastic matrix with partition

$$P = \begin{pmatrix} P_{11} & P_{12} \\ P_{21} & P_{22} \end{pmatrix}, \tag{418}$$

then the *stochastic complement* of P_{11} in P is defined as

$$S_{11} = P_{11} + P_{12} \left(I - P_{22}\right)^{-1} P_{21}. \tag{419}$$

Suppose we have two vectors $\begin{pmatrix} a \\ x \end{pmatrix}$ and $\begin{pmatrix} a \\ y \end{pmatrix}$, where a is a scalar, $\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$ is doubly stochastic matrix and

$$\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix} \begin{pmatrix} a \\ x \end{pmatrix} = \begin{pmatrix} a \\ y \end{pmatrix}, \tag{420}$$

so that $\begin{pmatrix} a \\ x \end{pmatrix}$ majorizes $\begin{pmatrix} a \\ y \end{pmatrix}$. Because

$$\begin{aligned} & \left[D_{22} + D_{21} \left(I - D_{11}\right)^{-1} D_{12}\right] x && (421) \\ = \;& D_{22}x + D_{21} \left(I - D_{11}\right)^{-1} \left(D_{12}x\right) && (422) \\ = \;& D_{22}x + D_{21} \left(I - D_{11}\right)^{-1} \left(I - D_{11}\right) a && (423) \\ = \;& D_{22}x + D_{21}a && (424) \\ = \;& y, && (425) \end{aligned}$$

pick $a = 1, x = y = e$ here, we know that $\begin{pmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{pmatrix}$ is doubly stochastic implies the stochastic complement $D_{22} + D_{21} \left(I - D_{11}\right)^{-1} D_{12}$ is doubly stochastic,

therefore x majorizes y. Multiplying a consecutive subblock product of elementary biunitary matrices to a vector acts as changing two elements in the vector once a time. If after every operation, we still have a majorization and we can continuing doing this without stop, then we will have a constructive proof of our extension.

Using the notation we have introduced, $\lambda \succ d$ tells us that $\lambda_1 \geq d_1$, and $\lambda_n \leq d_n$. We pick γ to be a permutation of λ and let $\gamma_n = \lambda_n$. At the first step, we choose

$$\gamma_{n-1} = \min \{\lambda_k \in \lambda : \lambda \geq d_n\}, \tag{426}$$

then we can find c_n and s_n, such that

$$d_n = e_2^T B_n \begin{pmatrix} \gamma_{n-1} \\ \gamma_n \end{pmatrix}. \tag{427}$$

For next $n-1$ steps, if

$$d_j > e_1^T B_{j+1} \begin{pmatrix} \gamma_j \\ \vdots \\ \gamma_n \end{pmatrix}, \tag{428}$$

we choose

$$\gamma_{j-1} = \min \{\lambda_k \in \lambda - \{\gamma_j, \ldots, \gamma_n\} : \lambda_k \geq d_j\}; \tag{429}$$

if

$$d_j \leq e_1^T B_{j+1} \begin{pmatrix} \gamma_j \\ \vdots \\ \gamma_n \end{pmatrix}, \tag{430}$$

then we choose

$$\gamma_{j-1} = \min \{\lambda_k \in \lambda - \{\gamma_j, \ldots, \gamma_n\}\}. \tag{431}$$

Thus we can find B_j such that

$$\begin{pmatrix} * \\ d_j \\ \vdots \\ d_n \end{pmatrix} = B_j \begin{pmatrix} \gamma_{j-1} \\ \gamma_j \\ \vdots \\ \gamma_n \end{pmatrix}, \tag{432}$$

continue the process, finally we have

$$\begin{pmatrix} * \\ d_2 \\ \vdots \\ d_n \end{pmatrix} = B_1 \begin{pmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{pmatrix}. \tag{433}$$

Since $d_1 + \cdots + d_n = \lambda_1 + \cdots + \lambda_n = \gamma_1 + \cdots + \gamma_n$, and $B_1$is biunitary, the first elemet in the left vector must be d_1. Therefore, $B = B_1$ and P is obtained from $\gamma = P\lambda$. We need to show that the process doesn't stop. For it to stop there must be either of the two cases.

Case 1. There exists $\lambda_p \leq d_k \leq \lambda_{p-1}$, such that

$$\begin{pmatrix} * \\ d_k \\ \vdots \\ d_n \end{pmatrix} = B_k \begin{pmatrix} \gamma_{k-1} \\ \gamma_k \\ \vdots \\ \gamma_n \end{pmatrix}, \tag{434}$$

where $\{\lambda_1, \ldots, \lambda_{p-1}\} \subset \{\gamma_{k-1}, \ldots, \gamma_n\}$, and $d_{k-1} > e_1^T B_k \begin{pmatrix} \gamma_{k-1} \\ \gamma_k \\ \vdots \\ \gamma_n \end{pmatrix}$.

But then

$$\begin{aligned} \sum_{i=1}^{n} d_i &= \sum_{i=1}^{k-2} d_i + \sum_{i=k-1}^{n} d_i && (435) \\ &> (k-2)\,\lambda_p + \sum_{i=k-1}^{n} \gamma_i && (436) \\ &\geq \sum_{i=1}^{n} \lambda_i, && (437) \end{aligned}$$

which contradicts the majorization.

Case 2. There exists $\lambda_p \leq d_k \leq \lambda_{p-1}$, such that

$$\begin{pmatrix} * \\ d_k \\ \vdots \\ d_n \end{pmatrix} = B_k \begin{pmatrix} \gamma_{k-1} \\ \gamma_k \\ \vdots \\ \gamma_n \end{pmatrix} \tag{438}$$

where $\{\lambda_1, \ldots, \lambda_{p-1}\} \subset \{\gamma_{k-1}, \ldots, \gamma_n\}$, and $d_{k-1} < e_1^T B_k \begin{pmatrix} \gamma_{k-1} \\ \gamma_k \\ \vdots \\ \gamma_n \end{pmatrix}$.

Since we always choose the smallest possible component of λ, we know that $\{\lambda_{k-1}, \ldots, \lambda_n\} = \{\gamma_{k-1,} \ldots, \gamma_n\}$ and therefore $\sum_{i=k-1}^{n} d_i < \sum_{i=k-1}^{n} \gamma_i$, which controdicts the majorization.

So the algorithm does not fail to continue.

A.2 Multi-period Quadratic Programming Solver with l^1 term

Let first solve the single period problem and then we will extend it to the multi-period case.

In the single period case, we are interested in the follwing optimization problem:

$$\min_w \frac{1}{2} w^T C^2 w - r^T w + c^T |w - w_0| \tag{439}$$

which is a quadratic plus positive L^1 term utility. The superscript T means "transpose" in linear algebra. For vectors a, b have the same length, $a^T b$ is defined as the inner product in the Euclidean space. To clarify, all vectors are column vectors instead of row vectors. You can also think C^2 as the covariance matrix whereas r as the vectors representing the direction to be pursued, we use C^2 to represent the covariance matrix because covariance matrices are positive definite, thus it would have a Cholesky decomposition. We say a matrix C^2 is positive definite if for any non-zero vector x, $x^T C^2 x > 0$. Factor model for the covariance matrix is also needed here, so C^2 has a diagonal plus low rank structure.

There would be two phases transformation to find an equivalent form of the optimization problem which is easier to solve. In phase 1, denote $w_* = w - w_0$ which could be regarded as the change of portfolio, plug it into (439),

$$\min_w \tfrac{1}{2} w^T C^2 w - r^T w + c^T |w - w_0| \tag{440}$$

$$\Leftrightarrow \quad \min_{w_*} \tfrac{1}{2}(w_0 + w_*)^T C^2 (w_0 + w_*) - r^T (w_0 + w_*) + c^T |w_*| \tag{441}$$

$$\Leftrightarrow \quad \min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r)^T w_* + c^T |w_*| + \left(\tfrac{1}{2} w_0^T C^2 w_0 - r^T w_0\right) \tag{442}$$

$$\Leftrightarrow \quad \min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r)^T w_* + c^T |w_*| \tag{443}$$

Let take a look at the L^1 term $c^T |w_*|$, suppose

$$c = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{pmatrix} \; and \; w_* = \begin{pmatrix} w_{*,1} \\ w_{*,2} \\ \vdots \\ w_{*,n} \end{pmatrix} \tag{444}$$

then

$$c^T |w_*| \tag{445}$$

$$= \sum_{i=1}^{n} c_i sign(w_{*,i}) w_{*,i} \tag{446}$$

$$= \sum_{i=1}^{n} \max_{t_i = c_i \; or \; -c_i} t_i w_{*,i} \tag{447}$$

$$= \max_{t_i = c_i \; or \; -c_i} t^T w_{*,i} \tag{448}$$

$$= \max_{-c \le t \le c} t^T w_* \tag{449}$$

The trick is plugging (449) in (443) and interchanging the min max order.

$$\min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r)^T w_* + c^T |w_*| \tag{450}$$

$$\Leftrightarrow \quad \min_{w_*} \left(\max_{-c \le t \le c} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r)^T w_* + t^T w_* \right) \tag{451}$$

$$\Leftrightarrow \quad \max_{-c \le t \le c} \left(\min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r + t)^T w_* \right) \tag{452}$$

Solving $\min_{w_*} \frac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r + t)^T w_*$ is trivial by taking derivative with

respect to w_*, immediately we have optimal $w_* = C^{-2}\left(r - C^2 w_0 - t\right)$ and

$$\begin{aligned} & \min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r + t)^T w_* && (453) \\ = \quad & -\tfrac{1}{2}\left(r - C^2 w_0 - t\right)^T C^{-2}\left(r - C^2 w_0 - t\right) && (454) \end{aligned}$$

Here C^{-2} stands for the inverse of C^2. Now we are only step to the phase one transformation, the orginal optimization problem is reduced to

$$\begin{aligned} & \max_{-c \le t \le c} \left(\min_{w_*} \tfrac{1}{2} w_*^T C^2 w_* + (C^2 w_0 - r + t)^T w_*\right) && (455) \\ \Leftrightarrow \quad & \max_{-c \le t \le c} \left(-\tfrac{1}{2}\left(r - C^2 w_0 - t\right)^T C^{-2}\left(r - C^2 w_0 - t\right)\right) && (456) \\ \Leftrightarrow \quad & \min_{-c \le t \le c} \tfrac{1}{2} t^T C^{-2} t - \left(C^{-2} r - w_0\right)^T t + \left[\left(r - C^2 w_0\right)^T C^{-2}\left(r - C^2 w_0\right)\right] && (457) \\ \Leftrightarrow \quad & \min_{-c \le t \le c} \tfrac{1}{2} t^T C^{-2} t - \left(C^{-2} r - w_0\right)^T t && (458) \end{aligned}$$

finally let us denote

$$f = -\left(C^{-2} r - w_0\right) \tag{459}$$

then our problem is equivalent to

$$\min_{-c \le t \le c} \frac{1}{2} t^T C^{-2} t + f^T t \tag{460}$$

The formula above is called the dual problem, thus we have successfully proved that the dual problem of a quadratic plus L^1 term utility is a box contraint quadratic problem. (Noticed that $-c \le t \le c$ is a box)

Before moving to phase 2, let briefly introduce the factor model, which has low rank structure of covariance matrix and is very commonly used [64, 65],

$$r = Vf + \epsilon \tag{461}$$

Suppose the return r (already demeaned) can be decomposed into a factor part Vf and idiosyncratic part ϵ, where V is called facor loading and f is called factor score. The assumption is f and ϵ are independent, f is mean zero with identity covariance($E(ff^T) = I$), ϵ is mean zero with diagonal covariance D

with posititive diagonal elements. ($D = E(\epsilon\epsilon^T)$) Then covariance for r is

$$\begin{aligned} & E(rr^T) && (462)\\ = \; & E(Vff^TV) + E(Vf\epsilon^T) + E(ef^TV^T) + E(\epsilon\epsilon^T) && (463)\\ = \; & VE(ff^T)V^T + E(\epsilon\epsilon^T) && (464)\\ = \; & VV^T + D && (465) \end{aligned}$$

A very important point everybody should keep in mind here is V is a tall and thin matrix, which means you only need a small amount of factors to explain the whole market. If $r \in \mathbb{R}^{n\times 1}$ and $V \in \mathbb{R}^{n\times p}$, then $n \gg p$. Let's say we have $3,000$ stocks, we may only need 20 factors. Since $rank(VV^T) = rank(V) \leq p$ and D is a diagonal matrix, now we have a diagonal plus low rank covariance matrix. For high dimensional covariance estimation, a legitimate thing to do is shrinkage, classical shrinkage estimator is just adding a contant times identity matrix to the sample covariance matrix, so it won't change our covariance structure because positive diagonal matrix plus mutiple of identity is still positive diagonal.

The other important point is the inverse of diagonal plus low rank matrix is diagonal minus low rank. Let start with one of the most commonly used matrix analysis formula *Woodbury matrix identity*:

$$(A+UCV)^{-1} = A^{-1} - A^{-1}U\left(C^{-1} + VA^{-1}U\right)^{-1}VA^{-1} \tag{466}$$

see wikipedia for more detail. By applying the formula,

$$\begin{aligned} & \left(D + VV^T\right)^{-1} && (467)\\ = \; & D^{-1} - D^{-1}V\left(I + V^TD^{-1}V\right)^{-1}V^TD^{-1} && (468) \end{aligned}$$

apparently $I + V^TD^{-1}V$ is positive definite and its Cholesky decomposition is $I + V^TD^{-1}V = LL^T$ where $L \in \mathbb{R}^{p\times p}$ is lower triangular. You may oberserved a simple fact: inverse of positive definite matrix is also positive definite. Now let

$$\begin{aligned} D_i &= D^{-1} && (469)\\ V_i &= D^{-1}VL^{-T} \in \mathbb{R}^{n\times p} && (470) \end{aligned}$$

then

$$\left(D + VV^T\right)^{-1} = D_i - V_iV_i^T \tag{471}$$

Let's go back to (460), first let $y = t + C^{-2}f$,

$$\begin{aligned} & \min_{-c \le t \le c} \frac{1}{2} t^T C^{-2} t + f^T t && (472) \\ \Leftrightarrow \quad & \min_{-c \le y - C^2 f \le c} \frac{1}{2} y^T C^{-2} y + \frac{1}{2} f^T C^2 f && (473) \\ \Leftrightarrow \quad & \min_{C^2 f - c \le y \le C^2 f + c} \frac{1}{2} y^T C^{-2} y && (474) \\ \Leftrightarrow \quad & \min_{C^2 f - c \le y \le C^2 f + c} \frac{1}{2} y^T \left(D_i - V_i V_i^T \right) y && (475) \end{aligned}$$

then let $x = D_i^{\frac{1}{2}} y$, since D has positive elements, $D_i^{\frac{1}{2}} \left(C^2 f - c \right) \le x \le D_i^{\frac{1}{2}} \left(C^2 f + c \right)$ is equivalent to $C^2 f - c \le y \le C^2 f + c$, we have to be careful this is not true if D has negative element. Also let

$$\begin{aligned} l &= D_i^{\frac{1}{2}} \left(C^2 f - c \right) && (476) \\ u &= D_i^{\frac{1}{2}} \left(C^2 f + c \right) && (477) \\ V_1 &= D_i^{-\frac{1}{2}} V_i && (478) \end{aligned}$$

finally the original optimization problem becomes

$$\begin{aligned} & \min_{C^2 f - c \le y \le C^2 f + c} \frac{1}{2} y^T \left(D_i - V_i V_i^T \right) y && (479) \\ \Leftrightarrow \quad & \min_{D_i^{\frac{1}{2}} (C^2 f - c) \le D_i^{\frac{1}{2}} y \le D_i^{\frac{1}{2}} (C^2 f + c)} \frac{1}{2} y^T D_i^{\frac{1}{2}} \left(I - D_i^{-\frac{1}{2}} V_i V_i^T D_i^{-\frac{1}{2}} \right) D_i^{\frac{1}{2}} y && (480) \\ \Leftrightarrow \quad & \min_{l \le x \le u} \frac{1}{2} x^T \left(I - V_1 V_1^T \right) x && (481) \end{aligned}$$

which is an incredible simple form, where V_1 is still a tall and thin matrix, $V_1 \in \mathbb{R}^{n \times p}$. Because of the positive definiteness of $D_i - V_i V_i^T$, $V_1 V_1^T$ is a contraction, meaning for any $x \in \mathbb{R}^n, \| V_i V_i^T x \| < \| x \|$, $I - V_1 V_1^T$ is a contraction as well. $l \le x \le u$ is a box contraint, the reason it's called box contraint is considering a two dimensional case, for example $l = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$ and $u = \begin{pmatrix} 2 \\ 4 \end{pmatrix}$, then the region $l \le x \le u$ is literally a box.

It turns out our algorithm for solving (481) is extremely simple:

$$\begin{aligned} x^{(0)} &= \min \left(\max \left(0, l \right), u \right) && (482) \\ x^{(k+1)} &= \min \left(\max \left(V_1 V_1^T x^{(k)}, l \right), u \right) && (483) \end{aligned}$$

Every iteration takes merely $O(np)$ operations and it always converges in less than 10 iterations among all the numerical simulation we conducted.

Now we provide the convergence analysis. In 1964, A.A.Goldstein published

a very interesting two-page paper [35].

Lemma 40. *In what follows P will denote the "projection" operator for the convex set C. This operator, which is well defined and Lipschitzian, assigns to given point in H its closest point in C. Take $x \in H$ and $y \in C$. Then $\langle x-y, P(x)-y\rangle \geq \|P(x)-y\|^2$.*

The inequality above is equivalent to $\langle P(x)-x, P(x)-y\rangle \leq 0$, $\langle\ ,\ \rangle$ stands for inner product. In our case it's just $(P(x)-x)^T (P(x)-y) \leq 0$. The lemma is quite intuitive, think C as an disk, then $P(x)$ is the tangential point, which implies the angle between $P(x)-x$ and any other vector $P(x)-y$ is larger than 90 degree.

A box is a convex set, define $P_{box}(x) = \min(\max(x,l),u)$, then obviously P_{box} is the "projection" will find the closest point onto the box. Denote the utility

$$u(x) = \frac{1}{2} x^T \left(I - V_1 V_1^T\right) x \tag{484}$$

then

$$\begin{aligned}
& u\left(P_{box}\left(V_1V_1^Tx\right)\right) - u(x) && (485)\\
= \ & \frac{1}{2} P_{box}\left(V_1V_1^Tx\right)^T \left(I - V_1V_1^T\right) P_{box}\left(V_1V_1^Tx\right) - \frac{1}{2} x^T \left(I - V_1V_1^T\right) x && (486)\\
= \ & x^T \left(I - V_1V_1^T\right)\left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (487)\\
& + \frac{1}{2}\left(P_{box}\left(V_1V_1^T\right) - x\right)^T \left(I - V_1V_1^T\right)\left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (488)\\
= \ & -\left(\left(V_1V_1^Tx\right) - x\right)^T \left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (489)\\
& + \frac{1}{2}\left(P_{box}\left(V_1V_1^T\right) - x\right)^T \left(I - V_1V_1^T\right)\left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (490)\\
\leq \ & -\left(P_{box}\left(V_1V_1^Tx\right) - x\right)^T \left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (491)\\
& + \frac{1}{2}\left(P_{box}\left(V_1V_1^T\right) - x\right)^T \left(I - V_1V_1^T\right)\left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (492)\\
= \ & -\frac{1}{2}\left(P_{box}\left(V_1V_1^T\right) - x\right)^T \left(I + V_1V_1^T\right)\left(P_{box}\left(V_1V_1^Tx\right) - x\right) && (493)
\end{aligned}$$

since $V_1V_1^T$ is contraction, $\frac{1}{2}\left(I + V_1V_1^T\right)$ is contraction as well, then there exists $c_0 \in (0,1)$ such that $u\left(P_{box}\left(V_1V_1^Tx\right)\right) - u(x) \leq c_0 \|P_{box}\left(V_1V_1^Tx\right) - x\|^2$. For the sequence getting from our algorithm,

$$u\left(x^{(k+1)}\right) - u\left(x^{(k)}\right) \leq -c\|x^{(k+1)} - x^{(k)}\|^2 \tag{494}$$

the utility decreases very fast, also u is continuous and has a lower bound 0, the algorithm has to converge.

Now let's extend the box quadratic programming algorithm to the multi-period case, we are interested in the problem,

$$\min_x \sum_{i=1}^{n} \left(\frac{1}{2} x_i^T H_i x_i - q_i^T x_i + c_i^T |x_i - x_{i-1}| \right), \tag{495}$$

where $H_i \in \mathbb{R}^{p \times p}$, which is equivalent to

$$\min_x \min_b \sum_{i=1}^{n} \left(\frac{1}{2} x_i^T H_i x_i - q_i^T x_i + b_i^T \left(x_i - x_{i-1} \right) \right) \tag{496}$$

subject to $-c_i \leq b_i \leq c_i, i = 1, 2, \ldots, n$. Let's define block shift matrix

$$Z = \begin{pmatrix} 0 & I & & \\ & \ddots & \ddots & \\ & & \ddots & I \\ & & & 0 \end{pmatrix}, \tag{497}$$

and block unit vector

$$E_1 = \begin{pmatrix} I \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{498}$$

Then the optimization problem becomes

$$\min_x \max_b \frac{1}{2} x^T H x - q^T x + b^T x - b^T \left(Z^T x \right) - \left(E_1^T b \right)^T x_0, \tag{499}$$

subject to $-c \leq b \leq c$. Because matrix multiplications are associative, we further have the optimization problem as

$$\min_x \min_b \frac{1}{2} x^T H x - \left(q - b + Zb \right)^T x - b^T \left(E_1 x_0 \right), \tag{500}$$

now let's switch the minimum and maximum sign, then

$$x_{opt} = H^{-1} \left(q - b + Zb \right). \tag{501}$$

By pluging x_{opt} back we obtain the dual optimization problem

$$\min_b \frac{1}{2} b^T \left[(I-Z)^T H^{-1} (I-Z) \right] b + b^T \left[E_1 x_0 - (I-Z)^T H^{-1} q \right] \tag{502}$$

subject to $-c \leq b \leq c$. Consider the structure of the new covariance matrix $(I-Z)^T H^{-1} (I-Z)$, first of all we have

$$H^{-1} = diag\left(H_k^{-1}\right) = diag\left(D_k - V_k V_k^T\right), \tag{503}$$

then

$$\begin{aligned}
& (I-Z)^T H^{-1} (I-Z) && (504)\\
= & \sum_{i=1}^{n-1} \begin{pmatrix} 0_{p(i-1)} & & & \\ & D_i - V_i V_i^T & -D_i + V_i V_i^T & \\ & -D_i + V_i V_i^T & D_i - V_i V_i^T & \\ & & & 0_{p(n-1-i)} \end{pmatrix} && (505)\\
& + \begin{pmatrix} 0_{p(n-1)} & \\ & D_n - V_n V_n^T \end{pmatrix} && (506)\\
= & \sum_{i=1}^{n} \begin{pmatrix} 0_{p(i-1)} & & & \\ & 2D_i & & \\ & & 2D_i & \\ & & & 0_{p(n-1-i)} \end{pmatrix} && (507)\\
& + \begin{pmatrix} 0_{p(n-1)} & \\ & 2D_n \end{pmatrix} && (508)\\
& - \sum_{i=1}^{n-1} \begin{pmatrix} 0_{p(i-1)} & & & \\ & D_i & D_i & \\ & D_i & D_i & \\ & & & 0_{p(n-1-i)} \end{pmatrix} && (509)\\
& - \sum_{i=1}^{n-1} \begin{pmatrix} 0_{p(i-1)} & & & \\ & V_i V_i^T & -V_i V_i^T & \\ & -V_i V_i^T & V_i V_i^T & \\ & & & 0_{p(n-1-i)} \end{pmatrix} && (510)\\
& - \begin{pmatrix} 0_{p(n-1)} & \\ & D_n + V_n V_n^T \end{pmatrix}. && (511)
\end{aligned}$$

The way we write down the formula makes it easy to see that $(I-Z)^T H^{-1} (I-Z)$

has the diagonal minus semi-positive definite structure, also the semi-positive definite part has low rank structure which makes the compuation very fast. Denote

$$
\begin{aligned}
f &= E_1 x_0 - (I - Z)^T H^{-1} q && (512)\\
&= \begin{pmatrix} x_0 - H_1^{-1} q_1 \\ H_1^{-1} q_1 - H_2^{-1} q_2 \\ \vdots \\ H_{n-1}^{-1} q_{n-1} - H_n^{-1} q_n \end{pmatrix}, && (513)
\end{aligned}
$$

similar to the single period case, we have the recursive algorithm,

$$
\begin{aligned}
& b_i^{k+1} && (514)\\
= \; & \min(c_i, \max(-c_i, (2D_{i-1} + 2D_i)^{-1} [&& (515)\\
& H_{i-1}^{-1} b_{i-1}^k + H_i^{-1} b_{i+1}^k - f_i && (516)\\
& + \left(D_i + D_{i-1} + V_{i-1} V_{i-1}^T + V_i V_i^T\right) b_i^k])) && (517)\\
= \; & \min(c_i, \max(-c_i, (2D_i + 2D_i)^{-1} [&& (518)\\
& V_{i-1} V_{i-1}^T \left(b_i^k - b_{i-1}^k\right) + V_i V_i^T \left(b_i^k - b_{i+1}^k\right) && (519)\\
& + D_{i-1} \left(b_i^k + b_{i-1}^k\right) + D_i \left(b_i^k + b_{i+1}^k\right) - f_i])), && (520)
\end{aligned}
$$

for $1 < i < p$. For the pivot case, we have

$$
\begin{aligned}
& b_1^{k+1} && (521)\\
= \; & \min(c_1, \max(-c_1, (2D_1)^{-1} [&& (522)\\
& V_1 V_1^T \left(b_1^k - b_2^k\right) + D_1 \left(b_1^k + b_2^k\right) - f_1])), && (523)
\end{aligned}
$$

and

$$
\begin{aligned}
& b_p^{k+1} && (524)\\
= \; & \min(c_p, \max(-c_p, (2D_p)^{-1} [&& (525)\\
& V_{p-1} V_{p-1}^T \left(b_p^k - b_{p-1}^k\right) + V_p V_p^T b_p^k && (526)\\
& + D_{p-1} \left(b_p^k + b_{p-1}^k\right) + D_p b_p^k - f_p])). && (527)
\end{aligned}
$$

References

[1] C.K.Chui, G.Chen, *Discrete H^∞ Optimization,* Springer Verlage, 1997

[2] R.A.Robert, C.T.Mullis, *Digital Signal Processing,* Addison-Wesley Series in Electrical Engineering

[3] A.P.Mullhaupt, K.S.Riedel, *Low Grade Matrices and Matrix Fraction Representations,* Linear Algebra and its Applications 342(2002), p.187-201

[4] G.Ammar, W.Gragg, L.Reichel, *Constructing a Unitary Hessenberg Matrix from Spectral Data,* Numerical Linear Algebra, Digital Signal Processing, and Parallel Algorithms, G.H. Golub and P. Van Dooren, eds., Springer-Verlag, Berlin, 1991, p.385-396

[5] P.A.Regalia, *Adaptive IIR Filtering in Signal Processing and Control,* CRC Press

[6] G.Szego, *Orthogonal Polynomials,* American Mathematical Society, 1967

[7] H.Dym, *Linear Algebra in Action,* American Mathematical Society, 2006

[8] A.Horn, *Doubly Stochastic Matrices and the Diagonal of a Rotation Matrix,* American Journal of Mathematics, VOL.36, no.3, 1954, p.620-630

[9] I.Bengtsson, A.Ericsson, M.Kus, W.Tadej, K.Zyczkowski, *Brikhoff's Polytope and Unistochastic Matrices, N=3 and N=4,* Commun. Math. Phys. 259, (2005), p.307-324

[10] K.Zyczkowski, W.Tadej, M.Kus, H-J.Sommers, *Random Unistochastic Matrices,* J. Phys. A:Math Gen 36 (2003) p.3425-3450

[11] B.Shao, *Approximation and Option Pricing for Fractional Brownian Motion*, Research note

[12] A.P.Mullhaupt, K.S.Riedel, *Band Matrix Representation of Triangular Input Balanced Form*, Research note

[13] A.P.Mullhaupt, K.S.Riedel, *Fast Adaptive Identification of Stable Innovation Filters,* IEEE Transactions on Signal Processing, VOL.45, no.10, 1997, p.2616-p2619

[14] A.P.Mullhaupt, *Triangular Input Balance, Meixner Functions, and a Nearly Rectangular Impulse Response*, Research note

[15] J.D.Makel, A.H.Gary, *Roundoff Noice Characterics of a Class of Orthogonal Polynomial Structures,* IEEE Transactions on Acoustics, Speech, and Signal Processing, VOL. ASSP-23, no.5, 1975, p.473-486

[16] J.D.Makel, A.H.Gary, *A Normalized Digital Filter Structure,* IEEE Transactions on Acoustics, Speech, and Signal Processing, VOL.ASSP-23, no.3, 1975, p.268 - 277

[17] P.A.Regalia, S.K.Mitra, P.P.Vaidyanathan, *The Digital All-Pass Filter: A Versatile Signal Processing Building Block,* Proceedings of the IEEE, VOL.76, no.1 January 1988, p.19 - 37

[18] B.Ninness, F.Gustafsson, *A Unifying Construction of Orthogonal Bases for System Identification,* IEEE Transaction on Automatic Control, VOL.42, no.4, p.515 - 521

[19] Peter.Heuberger, Paul.Van Den Hof, B.Wahlberg, *Modelling and Identification with Rational Orthogonal Basis Function,* Springer London

[20] R.G.Douglas, H.S.Shapiro, A.L.Shields, *Cyclic Vectors and Invariant Subspaces for the Backward Shift Operator,* Annales de l'institut Fourier, tome 20, no.1 (1970), p.37-76

[21] R.Peeters, M.Olivi, B.Hanzon, *Balanced Realization of Lossless Systems: Schur parameters, Canonical Forms and Applications,* 15th IFAC Symposium on System Identification, p.273-283

[22] B.Hanzon, M.Olivi, R.Peeters, *Balanced Realizations of Discrete-Time Stable All-Pass Systems and the Tangential Schur Algorithm,* Linear Algebra and its Applications, VOL.418, Issues.2-3, p.793-820

[23] L.Baratchart, M.Olivi, *Critical Points and Error Rank in Best H_2 Matrix Rational Approximation of Fixed McMillan Degree,* Constr. Approx.14, 1989, p.273-300

[24] J.Marmorat, M.Olivi, B.Hanzon, R.Peeters, *Matrix Rational H^2 Approximation: a State-Space Approach using Schur Parameters,* Proceedings of the 41st IEEE Conference on Decision and Control, VOL.4, p.4244 - 4249

[25] P.Fulcher, M.Olivi, *Matrix Rational H_2 Approximation: A Gradient Algorithm Based on Schur Analysis,* SIAM J. Control Optim, VOL.36, no.6, p.2013-2027

[26] A.Vandendorpe, *Model Reduction of Linear Systems, an Interpolation Point of View,* PhD Thesis

[27] R.Peeters, M.Olivi, B.Hanzon, *On a Recursive State-Space Method for Discrete-Time H_2-Approximation,* Research paper

[28] D.Alpay, L.Baratchart, A.Gombani, *On the Differential Structure of Matrix-Valued Rational Inner Functions,* Operator Theory: Advances and Applications VOL.73, 1994, pp 30-66

[29] X.Sun, *On Elementary Unitary and Φ-Unitary Transformations,* Research paper, 1995

[30] M.Olivi, *Parametrization of Rational Lossless Matrices with Applications to Linear System Theory*, PhD Thesis

[31] S.Amari, H.Nagaoka, *Methods of Information Geometry (Translations of Mathematical Monographs),* American Mathematical Society, 2007

[32] A.Ostrowski, *On Schur's Complement,* Journal of Combinatorial Theory, Series A, VOL.14, Issue.3, 1973, p.319–323

[33] D.Hinrichsen, A.J.Pritchard, *An Improved Error Estimate for Reduced-Order Models of Discrete-Time Systems,* IEEE Transaction on Automatic Control, VOL.35, no.3, p.317-320

[34] A.Bunse-Gerstner, D.Kubalinska, G.Vossen, D.Wilczek, *h_2-norm Optimal Model Reduction for Large Scale Discrete Dynamical MIMO systems,* Journal of Computational and Applied Mathematics 233, 2010 p.1202-1216

[35] A.A.Goldstein, *Convex programming in Hibert Space*, Bull. Amer. Math. Soc. VOL.70, no.5, 1964, p.709-710

[36] J.Decorte, A.Bultheel, M.Van Barel, *A Numerical Implementation of the Algorithm of Kung and Lin for AAK Model Reduction,* SIAM conference on Linear Algebra in Signals, Systems and Control, August, 1986, Boston

[37] A.Frazho, W.Bhosri, *An Operator Perspective on Signals and Systems,* Birkhäuser

[38] A.C.Antoulas, *Approximation of Large-Scale Dynamical Systems,* Society for Industrial and Applied Mathematic, 2005

[39] Y.M.Cho, G.Xu, T.Kailath, *Fast Identification of State-Space Models via Exploitation of Displacement Structure,* IEEE Transaction on Automatic Control, VOL. 39, no.10, 1994, p.45–59

[40] A.Sayed, T.Kailath, *A State-Space Approach to Adaptive RLS Filtering,* IEEE Signal Processing Magazine, p.1053-5888

[41] C.A.Beattie, S.Gugercin, *A Trust Region Method for Optimal $\mathcal{H}_2$ Model Reduction,* Joint 48th IEEE Conference on Decision and Control and 28th Chinese Control Conference, 2009, p.5370-5375

[42] K.Premaratne, E.Jury, M.Mansour, *An Algorithm for Model Reduction of 2-D Discrete Time Systems,* IEEE Transaction on Circuits and Systems, VOL.37, no.9, p.1116-1132

[43] S.Gugercin, *An Iterative SVD-Krylov based Method for Model Reduction of Large-scale Dynamical Systems,* Linear Algebra and its Application VOL.428, 2008, p.1964-1986

[44] C.Poussot-Vassal, *An Iterative SVD-Tangential Interpolation Method for Medium-Scale MIMO Systems Approximation with Application on Flexible Aircraft,* 2011 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC), p.7117 - 7122

[45] K.Abed-Meraim, W.Qiu, Y.Hua, *Blind System Identification,* Proceedings of the IEEE, VOL.85, no.8, p.1310-1322

[46] G.Flagg, C.Bettie, S.Gugercin, *Convergence of the Iterative Rational Krylov Algorithm,* Systems & Control Letters, VOL.61, no.6, 2012, p.688–691

[47] K.Kuo, P.Y.Wu, *Factorization of Matrices into Partial Isometries, Proceedings of The American Mathematical Society, VOL.105, no.2*

[48] S.Gugencin, A.C.Antoulas, C.Beattie, *$\mathcal{H}_2$ Model Reduction for Large-scale Linear Dynamical Systems,* SIAM J. Matrix Anal. Appl. VOL.30, no.2, pp. 609-638

[49] N.J.Young, *The Nevanlinna-Pick Problem for Matrix-Valued Functions,* J. Operator Theory, VOL.15, 1896, p.239-265

[50] C.Poussot-Vassal, P.Vuillemin, *Introduction to MORE: a MOdel REduction Toolbox,* 2012 IEEE International Conference on Control Applications, p.776-781

[51] E.Grimme, *Krylov Projection Methods for Model Reduction,* PhD Thesis

[52] H.Lev-Ari, T.Kailath, *Lattice Filter Parametrization and Modeling of Nonstationary Process,* IEEE Transaction on Information Theory, VOL.IT-30, no.1, p2-16

[53] E.Karlsson, M.Hayes, *Least Squares ARMA Modeling of Linear Time-Varying Systems: Lattice Filter Structures and Fast RLS Algorithms,* IEEE Transaction on Acoustics, Speech, and Signal Processing, VOL.ASSP-35, no.7, p.994-1014

[54] L.Baratchart, M.Olivi, F.Wielonsky, *On a Rational Approximation Problem in the Real Hardy Space H_2,* Theoretical Computer Science 94, 1992, p.175-197

[55] Y.Xu, T.Zeng, *Optimal $\mathcal{H}_2$ Model Reduction for Large Scale MIMO Systems via Tangential Interpolation,* International Journal of Numerical Analysis and Modeling, VOL.8, No.1, p.174-188

[56] S.Guttel, *Rational Krylov Approximation of Matrix Functions: Numerical Methods and Optimal Pole Selection,* GAMM-Mitteilungen, VOL.36, no.1, p.8–31

[57] G.Ammar, W.Gragg, *Schur Flows for Orthogonal Hessenberg Matrices,* Proceedings of the Fields Institute Workshop on Hamiltonian and Gradient Flows, Algorithms and Control, Waterloo, Canada, March, 1992

[58] S.Gugercin, J.R.Li, *Smith-Type Methods for Balanced Truncation of Large Sparse Systems,* Lecture Notes in Computational Science and Engineering VOL.45, 2005, p.49-82

[59] B.Ninness, H.Hjalmarsson, F.Gustafsson, *The Fundamental Role of General Orthonormal Bases in System Identification,* IEEE Transactions on Automatic Control, VOL.44, no.7, p.1384-1406

[60] G.Golub, C.Van Loan, *Matrix Computations,* Johns Hopkins University Press; fourth edition edition

[61] J.Choi, A.Mullhaupt, *Kahlerian Information Geometry for Signal Processing,* Research paper

[62] V.Peller, *An Excursion into the Theory of Hankel Operators,* Holomorphic Spaces MSRI Publications VOL.33, 1998

[63] R.Roshan, *Asymptotics for orthogonal polynomials, exponentially small perturbations and meromorphic continuations of Herglotz functions,* PhD Thesis

[64] D.B.Rubin, D.T.Thayer, *EM Algorithm for ML Factor Analysis, PYCHOMETRIKA-VOL.47. no.1, p.69-76*

[65] P.Diniz, *Adaptive Filtering: Algorithms and Practical Implementation,* Springer, 1997 edition