Matrix rational H^2 approximation: a state-space approach using Schur parameters

Jean-Paul Marmorat[1], Martine Olivi[2], Bernard Hanzon[3], Ralf L.M. Peeters [4]

Abstract. This paper deals with the problem of computing a best *stable* rational L^2 approximation of specified order to a given *multivariable* transfer function. The problem is equivalently formulated as a minimization problem over the manifold of *stable all-pass* (or lossless) transfer functions of fixed order [6]. Some special Schur parameters are used to describe this manifold [10]. Such a description presents numerous advantages: it takes into account the stability constraint, possesses a good numerical behavior and provides a model in state-space form, which is very useful in practice. A rigorous and convergent algorithm is proposed to compute local minima which has been implemented using standard MATLAB subroutines. The effectiveness of our approach to solve model reduction problems as well as identification problems in frequency domain is demonstrated through several examples, including real-data simulations.

1 Introduction

The identification of linear time-invariant systems can be formulated as a rational approximation problem in which some cost function is optimized over a set of systems. In this paper, we consider discrete-time, causal systems represented by their proper transfer functions. The cost function is the L^2 norm, so that our approximation problem states in the space $\bar{H}_2^{m\times p}$, where $\bar{H}_2$ denotes the complex Hardy space of the complement of the unit disk (see [8]), as follows:

[1]CMA, BP 93, 06902 Sophia-Antipolis Cedex, FRANCE, {marmorat}@sophia.inria.fr, phone: 33 4 92 38 79 56, fax: 33 4 92 38 79 98

[2]INRIA, BP 93, 06902 Sophia-Antipolis Cedex, FRANCE, {olivi}@sophia.inria.fr, phone: 33 4 92 38 78 77, fax: 33 4 92 38 78 58

[3]Dept. Econometrics, FEWEC, Vrije Universiteit, De Boelelaan 1105, 1081 HV Amsterdam, The Netherlands, Tel.: +31-20-446017, Fax: +31-20-44446020, Email: bhanzon@econ.vu.nl

[4]Dept. Mathematics, Universiteit Maastricht, P.O. Box 616, 6200 MD Maastricht, The Netherlands, Tel.: +31-43-3883365, Fax: +31-43-3884910, Email: ralf.peeters@math.unimaas.nl

(RAP) *given a transfer function $F \in \bar{H}_2^{m\times p}$, minimize*

$$\|F-H\|_2^2 = \frac{1}{2\pi}\mathrm{Tr}\int_0^{2\pi}[F-H](e^{it})[F-H](e^{it})^*dt,$$

as H ranges over the set of rational stable (i.e. analytic for $|z|>1$) functions of McMillan degree n.

It was proved [2] that *the global minimum of the L^2 criterion does exist.*

First remark that if H is a solution to (RAP), we must have $H(\infty)=F(\infty)$. Thus, we may restrict our study to the case of strictly proper transfer functions. We denote by $\bar{H}_{2,0}$ the orthogonal complement of the Hardy space H^2 of the unit disk, or equivalently the subspace of $\bar{H}_2$ consisting with strictly proper transfer functions. A strictly proper transfer function can be represented by means of the (right) Douglas–Shapiro–Shields factorization (see. [5]):

$$H \;=\; P\,G, \tag{1}$$

where G is a $(p\times p)$-rational lossless function, P is a $(m\times p)$-rational unstable matrix, and G and H have same McMillan degree. The lossless factor is unique up to a right unitary factor. The set of $\mathbb{C}^{p\times p}$-valued rational lossless functions of degree n will be denoted by $\mathcal{L}_n^p$, and the set of unitary $p\times p$ constant matrices by $\mathcal{U}_p$. We shall say that two members G_1 and G_2 of $\mathcal{L}_n^p$ are equivalent if $G_1 = U\,G_2$, for some $U\in\mathcal{U}_p$. The associated quotient space will be denoted by $\mathcal{L}_n^p/\mathcal{U}_p$. Note that the function H in (1) belongs to the set $\mathcal{H}(G)$, the orthogonal complement of $H_2^{m\times p}\,G$ into $\bar{H}_{2,0}^{m\times p}$. Since $\mathcal{H}(G)$ is a vector space, (RAP) *is equivalent to* the minimization of the function

$$\begin{array}{cccc}\Psi^n & : & \mathcal{L}_n^p/\mathcal{U}_p & \rightarrow & \mathbb{R}\\ & & G & \rightarrow & \|F-\pi_n(G)\|_2^2,\end{array} \tag{2}$$

where $\pi_n(G)$ denotes the projection of F onto $\mathcal{H}(G)$. It was proved in [1] that $\mathcal{L}_n^p$ and its quotient

$\mathcal{L}_n^p/\mathcal{U}_n^p$ are smooth manifolds of dimension $2np+p^2$ and $2np$ respectively. 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 $\mathcal{L}_n^p/\mathcal{U}_p$ is trivial (it is an open subset of a Euclidean space) and an algorithm to find the local minima of (2) is described in [3]. In [6], these results are extended to the multivariable case. The main difficulty was to find a nice parametrization for the manifold $\mathcal{L}_n^p/\mathcal{U}_n^p$. It was done by means of Schur parameters, provided by a tangential Schur algorithm [1], and used as local coordinates. The optimization problem is then tackled 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. It must be noted that in this approach, *the stability constraint is directly taken into account by the parametrization.*

In this paper, we follow the same approach but we make a different choice in the tangential Schur algorithm which provides a state-space representation for the lossless functions [10]. This parametrization allows to express the criterion directly in state-space form and presents a very nice numerical behavior. It combines the technical advantages of the Schur parametrization to the practical ones of the state-space description. A gradient type algorithm has been implemented in MATLAB and some results are presented here, including real-data simulations coming from an hyperfrequency filter.

The natural framework for our study is the complex case, that is the case of functions whose Fourier coefficients are complex. The real case, which is relevant in most applications, can be handle by this method (see Examples 1 and 2). However a specific treatment will be relevant and is under study.

2 Parametrization of lossless functions

Let G be a $(p \times p)$-lossless function of degree n. Such a function admits a balanced realization

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

such that the associated *realization matrix*

$$R = \begin{bmatrix} D & C \\ B & A \end{bmatrix} \qquad (3)$$

is *unitary* (see [10] and the bibliography therein).

The following points are essential for our purpose (see [6] and [10] for more details and for the proofs). Let

$$G(1/\bar{w})u \;=\; v, \qquad (4)$$

$w \in \mathbb{C}, |w| < 1$, $u \in \mathbb{C}^p$, $\|u\| = 1$ and $v \in \mathbb{C}^p$, $\|v\| \leq 1$, be some interpolation condition:

(i) $G(z)$ and w being given, we can always find some direction u such that v given by (4) has norm *strictly less* than 1.

(ii) in that case ($\|v\| < 1$), a $(p \times p)$ block-matrix function

$$\Theta(z) \;=\; \begin{pmatrix} \Theta_1(z) & \Theta_2(z) \\ \Theta_3(z) & \Theta_4(z) \end{pmatrix} \qquad (5)$$

can be defined, depending on w, u, v, such that the function $G(z)$ can be represented by the linear fractional transformation

$$G \;=\; (\Theta_4\, G_{n-1} + \Theta_3)\,(\Theta_2\, G_{n-1} + \Theta_1)^{-1}, \qquad (6)$$

for some *lossless function $G_{n-1}(z)$ of degree $n-1$*. The (J-unitary) matrix function $\Theta(z)$ is uniquely determined up to a constant (J-unitary) right factor H.

(iii) for a particular choice of the factor H in $\Theta(z)$, a very simple transformation on realizations corresponds to the linear fractional transformation (6): let R_{n-1} be a unitary realization matrix of $G_{n-1}(z)$, then a unitary realization matrix R_n of $G(z)$ is given by

$$R_n = \begin{bmatrix} V & 0 \\ 0 & I_n \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & R_{n-1} \end{bmatrix} \begin{bmatrix} U^* & 0 \\ 0 & I_n \end{bmatrix}, \qquad (7)$$

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

$$U \;=\; \begin{bmatrix} \xi u & I_p - (1+w\eta)uu^* \\ \overline{w}\eta & \xi u^* \end{bmatrix},$$

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

with

$$\xi = \frac{\sqrt{1-|w|^2}}{\sqrt{1-|w|^2\|v\|^2}}, \qquad \eta = \frac{\sqrt{1-\|v\|^2}}{\sqrt{1-|w|^2\|v\|^2}}.$$

The tangential Schur algorithm consists in repeating this process, and thus provides a sequence of lossless functions $G_k(z)$ of degree k , satisfying the interpolation condition

$$G_k(1/\bar{w}_k)u_k = v_k, \quad \|v_k\| < 1,$$

until G_0 which is a *constant unitary matrix.*

As in the scalar case, the interpolation values $v_n = v, v_{n-1}, \ldots, v_1$, can be taken as parameters to describe the quotient space $\mathcal{L}_n^p/\mathcal{U}_p$. But in the matrix case, they only describe an open subset of the manifold. Associated with the sequences

$$\begin{aligned}\mathbf{w} &= (w_n = w, w_{n-1}, \ldots, w_1),\\ \mathbf{u} &= (u_n = u, u_{n-1}, \ldots, u_1),\end{aligned}$$

of interpolation points and interpolation directions, and with an unitary $(p \times p)$ matrix D_0, we define a chart $(\mathcal{V}, \varphi)$ by its domain

$$\mathcal{V}_{(\mathbf{w},\mathbf{u},D_0)} = \{G \in \mathcal{L}_n^p \;/\; \|G_k(1/\bar{w}_k)u_k\| < 1, G_0 = D_0\},$$

and its coordinate map :

$$\varphi \;:\; G \rightarrow (v_1, v_2, \ldots, v_n).$$

The family $(\mathcal{V}, \varphi)$ defines a C^∞ atlas on $\mathcal{L}_n^p$.

In such a chart, a *unitary realization matrix* of the current lossless function G can be computed by iterating formula (7), which presents a very nice numerical behavior since it only involves *multiplications by unitary matrices.*

3 State-space formulas for the criterion.

It is well-known (see [1]) that the manifold $\mathcal{L}_n^p/\mathcal{U}_p$ is diffeomorphic to the set $\mathcal{R}_p^-(n)$ of reachable pairs (A, B) for which the spectrum of A is in the open unit disk, quotiented by the usual equivalence relation. Let

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

be a minimal realization of a lossless function $G(z) \in \mathcal{L}_n^p$ such that the associated realization matrix (3) is unitary.

We now give formulas for the L^2 criterion $\Psi_n(G)$, in function of the reachable pair (A, B). First remark that the projection $H = \pi_n(G)$ must have a realization of the form

$$H(z) = \gamma(zI_n - A)^{-1}B. \qquad (8)$$

Depending on the application we have in mind, the function F can be given

(1) by a realization

$$F(z) = \mathcal{C}(zI_N - \mathcal{A})^{-1}\mathcal{B}. \qquad (9)$$

In this case, the error $F - H$ has realization

$$A_g = \begin{bmatrix} \mathcal{A} & 0 \\ 0 & A \end{bmatrix},$$

$$B_g = \begin{bmatrix} \mathcal{B} \\ B \end{bmatrix}, \quad C_g = \begin{bmatrix} \mathcal{C} & -\gamma \end{bmatrix}.$$

Let W_g be the solution to the Lyapunov equation

$$A_g^* W_g A_g + C_g^* C_g = W_g.$$

It is well-known that the L^2 norm is given by

$$\|F - H\|_2^2 = \operatorname{Tr}\left(B_g^* W_g B_g\right).$$

Let W_g be partitioned accordingly to the partitioning of A_g:

$$W_g = \begin{bmatrix} W_{11} & W_{12} \\ W_{21} & I_n \end{bmatrix},$$

(the reachability gramian $W_{2,2}$ of the pair (A, B) is the identity matrix since the realization matrix is unitary). It can be block-diagonalized as follows

$$W_g = \begin{bmatrix} I_N & W_{12} \\ 0 & I_n \end{bmatrix} \Delta \begin{bmatrix} I_N & 0 \\ W_{21} & I_n \end{bmatrix},$$

with

$$\Delta = \begin{bmatrix} W_{11} - W_{12}W_{21} & 0 \\ 0 & I_n \end{bmatrix}$$

The pair (A, B) being given, the error norm is thus minimal for

$$\gamma = \mathcal{C}W_{12} \qquad (10)$$

and the criterion given by

$$\Psi_n(G(z)) = \operatorname{Tr}(\mathcal{C}W_{11}\mathcal{C}^*) - \operatorname{Tr}(\gamma\gamma^*). \qquad (11)$$

(2) by its Fourier coefficients up to some order K:

$$F(z) = \sum_{k=1}^{K} F_k z^{-k}.$$

Note that in practice, F has finite (high) order. Let $(\mathcal{A}, \mathcal{B}, \mathcal{C})$ be some realization of F. Then, the computation in (1) is still valid and $F_k = \mathcal{C}\mathcal{A}^{k-1}\mathcal{B}$. Moreover W_{12} is given by

$$W_{12} = \sum_{k \geq 0} \mathcal{A}^k \mathcal{B} B^* (A^*)^k,$$

and (10) can be rewritten

$$\gamma = \sum_{k\geq 0} F_{k+1} B^*(A^*)^k, \qquad (12)$$

so that

$$\Psi_n(G(z)) = \sum_{k=1}^{K} \mathrm{Tr}(F_k^* F_k) - \mathrm{Tr}\,(\gamma\gamma^*)\,. \qquad (13)$$

4 The algorithm

The algorithm receives as inputs

- The $(m\times p)$-matrix function F to be approximate that can be given in one of the forms
  - **(i)** a realization $(\mathcal{A},\mathcal{B},\mathcal{C},\mathcal{D})$ of order N,
  - **(ii)** its $K+1$ first Fourier coefficients $F_0, F_1, F_2, \ldots, F_K$,
- the degree n of the approximant,
- an initial $(p\times p)$-lossless function of degree n given by some unitary realization (it may be supplied by a random process).

The algorithm proceeds as follows: an adapted chart (that is, as explained in section 2, a sequence of interpolation points and interpolation conditions) is constructed for the given initial lossless functions (see case A. given below). A minimization process is started which minimizes the local criterion $\psi_n(v_1,\ldots,v_n)$, where $(v_1,\ldots,v_n)$ are the Schur parameters in the chart, subject to the constraints $\|v_i\| < 1,\ i = 1\ldots n$. The current lossless function is computed from the Schur parameters using recursively formula (7) which provides a unitary realization (see section 2) and then, the criterion is computed from the reachable pair via (11) or (13) depending on the form of the data (see section 3). If the norm of some Schur parameter tends to 1 (see case B. below), either the chart is no more available and we must find a new adapted chart, or we have reached a boundary point of the manifold, of degree less that the current degree, say k. This case is unusual, and if it does not happen, at the end a local minima is reached.

Remark. In practice we proceed in that way when $p\leq m$, while when $p>m$ we preferably work with the transpose of the given function, in order to have less parameters to minimize. Indeed, when we approximate an $(m\times p)$-function, the number of parameters is equal to $2np$ while when approximate its transpose the number of parameters is equal to $2nm$.

A. Choice of an adapted chart.
The present version of our algorithm uses a selection strategy which only works in the complex case. Observe that if we choose a Schur parameter equal to zero, recursion (7) applied to the unitary realization matrix (3) gives

$$\left[\begin{array}{c|cc} \hat{D} & \sqrt{1-|w|^2}Du & C \\ \hline \sqrt{1-|w|^2}u^* & w & 0 \\ B(I_p-(1+\bar{w})uu^*) & \sqrt{1-|w|^2}Bu & A \end{array}\right]$$

where $\hat{D} = D(I_p-(1+\bar{w})uu^*)$. This observation suggests the following strategy which leads to a chart in which all the Schur parameters are equal to the null vector and corresponds to a Potapov factorization [12]. Starting from any balanced realization of the lossless function $G(z)\in\mathcal{L}_n^p$, compute a new realization (A_n,B_n,C_n,D_n), still balanced but in which A_n is lower triangular (Schur form). Let

$$A_n = \left[\begin{array}{cc} w_n & 0\cdots 0 \\ \beta_n & A_{n-1}\end{array}\right], \quad B_n = \left[\begin{array}{c} b_n \\ \vdots \end{array}\right],$$

$$C_n = \left[\begin{array}{cc} c_n & C_{n-1}\end{array}\right],$$

where w_n is a complex number, b_n a row vector of size p, and c_n a column vector of size p. Put

$$u_n = b_n^*/\|b_n\|,$$

and

$$\begin{aligned} B_{n-1} &= B_n + \frac{1+w_n}{1-|w_n|^2}\beta_n b_n^*, \\ D_{n-1} &= D_n + \frac{1+w_n}{1-|w_n|^2}c_n b_n^*. \end{aligned}$$

Repeating this process yields a sequence of complex points of the open unit disk $(w_n,\ldots,w_1)$, a sequence of unit vectors $(u_n,\ldots,u_1)$ and a unitary matrix D_0 that index a chart in which G has Schur parameters $v_n=\ldots=v_1=0$.

B. The norm of some Schur parameter tends to 1.
First remark that if the vector v as norm 1, formula (7) keeps sense, and applied to the unitary realization matrix (3) gives the unitary realization matrix

$$\left[\begin{array}{c|cc} \hat{D} & (I_p-vv^*)Du & (I_p-vv^*)C \\ \hline -v^*D(I_p-uu^*) & -v^*Du & -v^*C \\ B(I_p-uu^*) & Bu & A \end{array}\right]$$

where $\hat{D} = vu^* + (I_p - vv^*)D(I_p - uu^*)$.
Assume that $\|v_i\| = 1$, then,
- either A_{i+1} has some eigenvalue of norm one, the realization is not minimal and the lossless function $G(z) = \varphi^{-1}_{(\mathbf{w},\mathbf{u},D_0)}(v_1, \ldots, v_n)$ has degree less than the current degree, say degree k. This point belongs to the boundary of the manifold. In that case, we stop the procedure.
- either all the eigenvalues of A_{i+1} have norm less than one and we must change the chart (see case A. above) before restarting the minimization procedure.

The main drawback of the L^2 norm is that it possesses many local minima while our final goal is to find the global one. However, we can think of using the result of a balanced truncation method, for example, as a starting point for our minimization procedure. Another strategy is under study [11] that could warrant a rapid convergence to the global minimum. It consists in an iterative search on the degree (see [6]), a minimum at order k providing a starting point for the minimization procedure at degree $k+1$, which ensures an improvement of the criterion as the degree increases. It also allows to restart the minimization process when a boundary point of degree less than the current degree is reached (see case B. above).

5 Application to continuous-time systems

This approximation procedure applies to continuous-time systems by means of an isometry between them and discrete-time systems. Continuous-time systems that we consider are those whose transfer functions belong to the Hardy space $\mathcal{H}^2_{m\times p}$ of the right-half plane endowed with the L^2-norm

$$\|\tilde{F}\|_c = \frac{1}{2\pi}\mathrm{Tr} \int_{-\infty}^{\infty} F(iy)F(iy)^*\,dy. \qquad (14)$$

Then

$$F(z) = \frac{1}{z-1}\tilde{F}\left(\frac{z+1}{z-1}\right) \qquad (15)$$

belongs to $H^{m\times p}_{2,0}$ and we have $\|\tilde{F}\|^2_c = 2\|F\|^2_2$. If $\tilde{F}(s) = \tilde{C}(sI_n - \tilde{A})^{-1}\tilde{B}$ is a realization of $\tilde{F}$, then $F(z) = C(zI_n - A)^{-1}B$ is a realization of $F(z)$, with

$$\begin{aligned} C &= \tilde{C}, \\ A &= -(I_n - \tilde{A})^{-1}(I_n + \tilde{A}), \\ B &= (I_n - \tilde{A})^{-1}\tilde{B}. \end{aligned} \qquad (16)$$

This process can be inverted proving that the isometry also *preserves McMillan degree.*

6 Numerical Examples

The software MATLAB was used for the implementation, and in particular the function *fmincon* in the optimization stage and the function *schur* in the construction of an adapted chart. The software is named RARL2.

Example 1. We consider a MIMO model reduction problem in continuous time. It is an automobile gas turbine model with 2 inputs, 2 outputs and 12 states, given by a realization in [9, p.168]. It was also studied in [7] and [13] among others. In table 1, our results are summarized and compared with the results of [13] and with the results obtained by balanced truncation (BT). A very good approximant is obtained at degree 8, as expected from the computation of the Hankel singular values [7]. Note that the other methods fail to give it.

Table 1. The comparison of the relative errors among three methods.

Deg	RARL2	[13]	BT
4	0.13481	0.1354	0.3687
5	0.07799	0.0795	0.1295
6	0.05248	0.0541	0.1151
8	$8.64\ 10^{-4}$	-	-

Remark. In this example, the system is real. Even if our procedure is ignorant of this information, starting from a real initial system we found almost real (imaginary part less than 10^{-3}) approximants.

Example 2. The problem is to find a 8th order model of a MIMO ((2×2)) hyperfrequency filter, from experimental pointwise frequency values provided by the CNES (French space agency). A first stage, far from being trivial, consists in computing a stable matrix transfer function of high order which represents the filter. It is achieved by the software Hyperion, also developed at INRIA [4]. This function is given by its 800 first Fourier coefficients and is then approximated by RARL2. Very good results are obtained. They are plotted on figure 1.

Example 3. This example shows that a "real" system may have a complex global minimum. It is

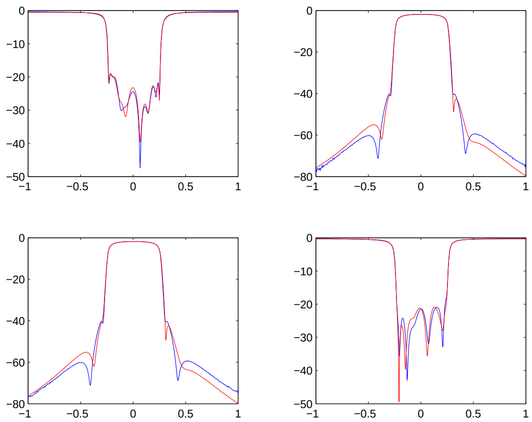

Figure 1: CNES (2×2) hyperfrequency filter: data and approximant at order 8 (Bode diagram).

the discrete-time system whose transfer function is

$$f(z) = \frac{1 - z^2}{z^3}.$$

The $L2$ criterion at order 1 admits tree minima: a real and two complex ones symmetric with respect to the real axis (see figure 2). It happens that the relative error at the real minimum is 0.7071068 while at both complex ones it is 0.6382847, so that there are two complex best approximants in that case.

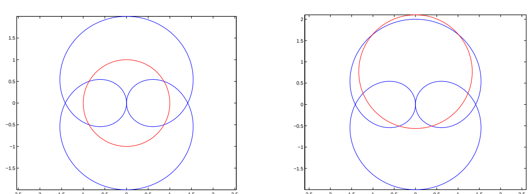

Figure 2: Nyquist diagram of example 4, one of its complex approximants (on the right) and its real approximant (on the left).

References

[1] D. Alpay, L. Baratchart, and A. Gombani, *On the differential structure of matrix-valued rational inner functions*, Operator Theory : Advances and Applications, 73 (1994), pp. 30–66.

[2] L. Baratchart, *Existence and generic properties for L^2 approximants of linear systems*, I.M.A. Journal of Math. Control and Identification, 3 (1986), pp. 89–101.

[3] L. Baratchart, M. Cardelli, and M. Olivi, *Identification and rational L^2 approximation : a gradient algorithm*, Automatica, 27, No.2 (1991), pp. 413–418.

[4] L. Baratchart, J. Grimm, J. Leblond, M. Olivi, F. Seyfert, and F. Wielonsky, *Identification d'un filtre hyperfréquence.* Rapport Technique INRIA No 219.

[5] R. Douglas, H. Shapiro, and A. Shields, *Cyclic vectors and invariant subspaces for the backward shift operator*, Annales de l'Institut Fourier (Grenoble), 20 (1970), pp. 37–76.

[6] P. Fulcheri and M. Olivi, *Matrix Rational H^2-Approximation: a Gradient Algorithm Based on Schur Analysis*, SIAM J. Contr. and Opt., 36 (1998), pp. 2103–2127.

[7] K. Glover, *All optimal Hankel-norm approximations of linear multivariable systems and their L^∞-error bounds*,Int. J. Control 39 (1984), pp. 1115-1193.

[8] K. Hoffman, *Banach spaces of analytic functions*, Dover, 1988.

[9] Y.S. Hung and A.G.J. MacFarlane, *Multivariable feedback: a quasi-classical approach*, Lect. Notes in Control Sci. 40, Springer- Verlag, 1982.

[10] R.L.M. Peeters, B. Hanzon and M. Olivi, *Balanced realizations of discrete-time stable all-pass systems and the tangential Schur algorithm*, in Proceedings of the ECC99, Karlsruhe, Germany, August 31-September 3.

[11] R.L.M. Peeters, B. Hanzon and M. Olivi, *On a recursive state-space method for discrete-time H_2-approximation*, in Proceedings of the MTNS02, Notre Dame (Indiana), USA, August 12-16.

[12] V.P. Potapov, *The multiplicative structure of J-contractive matrix functions*, Amer. Math. Soc. Transl. (2) 15 (1960), pp. 131–243 .

[13] W.-Y. Yan and J. Lam, *An approximate approach to H^2 optimal model reduction*, IEEE Trans. Autom. Control 44, No.7 (1999), pp 1341-1358.