Identification and matrix rational H^2 approximation : a gradient algorithm based on Schur analysis

Pascale Fulcheri, Martine Olivi

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Identification and matrix rational H^2 approximation : a gradient algorithm based on Schur analysis.

Pascale Fulcheri, Martine Olivi

N° 2520

Avril 1995

PROGRAMME 5

Rapport de recherche

ISSN 0249-6399

Identification and matrix rational H^2 approximation : a gradient algorithm based on Schur analysis.

Pascale Fulcheri, Martine Olivi

Programme 5 — Traitement du signal, automatique et productique
Projet MIAOU

Rapport de recherche n° 2520 — Avril 1995 — 31 pages

Abstract: This report deals with rational approximation of any specified order n of transfer functions. Transfer functions are assumed to be matrices whose entries belong to the Hardy space for the complement of the closed unit disk endowed with the L^2-norm. A new approach is developed leading to an original algorithm, the first one to our knowledge which concerns matrix transfer functions. This approach generalizes the ideas developed in the scalar case, but involves substantial new difficulties. The inner-unstable factorization of transfer functions allows to express the criterion in terms of inner matrices of Mac-Millan degree n. These matrices form a differential manifold. Based on a tangential Schur algorithm, an atlas of this manifold is given for which the coordinates vary in n copies of the unit ball. Then a gradient algorithm can be used to solve this problem. The different cases which can arise while processing the algorithm are studied : how to switch to another chart of the atlas, what has to be done when a boundary point is reached. In the neighbourhood of a boundary point, the criterion can be smoothly extended. Moreover, such a point can be considered as an initial point for the research of a lower degree approximant. It is explained how to cope with the decrease and the increase of the degree. The convergence of the algorithm to a local minimum of appropriate degree is proved and demonstrated on a simple example.

Key-words: rational approximation, identification, discret time systems, inner matrices, gradient algorithm, Schur analysis.

(Résumé : tsvp)

Unité de recherche INRIA Sophia-Antipolis
2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex (France)
Téléphone : (33) 93 65 77 77 – Télécopie : (33) 93 65 77 65

Identification
et approximation rationnelle matricielle dans H^2 :
un algorithme du gradient basé sur l'analyse de Schur.

Résumé : Ce rapport traite de l'approximation rationnelle en norme L^2, à degré fixé, des fonctions de transfert. Une nouvelle approche est proposée qui conduit à un algorithme original, le premier algorithme matriciel à notre connaissance. Cette approche généralise les idées développées dans le cas scalaire, mais nécessite la mise en oeuvre de nouveaux outils mathématiques. La factorisation intérieure-instable des fonctions de transfert stables permet d'exprimer le critère en fonction des matrices intérieures de degré fixé. Il s'agit alors de paramétrer l'ensemble de ces matrices et d'en étudier finement la structure. En utilisant la théorie des espaces à noyaux reproduisants, on montre qu'il s'agit d'une variété différentielle dont on construit un atlas. Dans une carte, une matrice intérieure de degré n est caractérisée par n vecteurs de norme strictement inférieure à 1 appelés paramètres de Schur. On est alors en mesure de développer un algorithme du gradient permettant de trouver les minimums locaux. Le problème majeur posé par cet algorithme réside dans le fait qu'il peut générer des paramètres de norme 1. L'étude des propriétés du critère au voisinage de tels paramètres montre alors comment adapter l'algorithme pour garantir sa convergence.

Mots-clé : identification, approximation rationnelle, systèmes discrets, matrices intérieures, espaces à noyaux reproduisants, analyse de Schur, algorithme du gradient.

1 Introduction.

This work takes place in the framework of *model reduction* : this process consists in replacing an elaborate model of a physical system by a simpler one which is easier to analyse and to use, without incurring too much error.

Our systems will be assumed to be linear, time invariant and causal and thus admit a transfer function (cf. [7]). A transfer function is a matrix of power series which applied to an input (represented by a vector of power series) gives the corresponding output. A good measure of the complexity of a system is its *order*, that is the number of state variables in a minimal realization. The transfer function of a finite order system is a matrix of rational functions, and this order can be computed as the Mac-Millan degree of the transfer function or equivalently its number of poles counted with their multiplicities (see [9]).

We can state our problem of model reduction in terms of transfer functions : given a transfer function, find another transfer function of significant lower degree with comparable performance. Now to give the problem a mathematical content, we need a measure of performance. This will be done by *rational approximation*, requiring our transfer function to lie in a Banach space.

We choose to work in discret time. In this context, a rational system will be stable, i. e. a bounded input produces a bounded output, if its transfer function T has analytic entries outside the unit disk $\mathbf{U}$, or else if the poles of T range over the closed unit disk. We also assume that the transfer function is strictly proper with no lost of generality. Let H^p_-, $1 \leq p < \infty$ be the Hardy space of functions f analytic outside the closed unit disk vanishing at infinity, which satisfy

$$\|f\|_p = sup_{r>1} \int_{-\pi}^{\pi} |f(re^{it})|^p dt < \infty.$$

The entries of T, being rational stable and strictly proper, belong to any Hardy space H^p_-, for $1 \leq p < \infty$, and each corresponding norm provides a measure of performance. Note that the results can be carried over to continuous time by considering Hardy spaces of the right half plane (cf. [8]).

In this work, the criterion which is chosen is the L^2-norm. The L^2-approximation problem can be stated as follows : given F, a $p \times m$ stable transfer function, find a rational stable function H of degree at most n which minimizes the L^2-norm of $F - H$. A lot of qualitative results have been proved in [2], and in particular the following property : if F is not itself rational of degree at most n, then a best local approximant H has degree exactly n. Thus, we can restrict our search to the smooth manifold of stable systems of degree n. The L^2 criterion is a good mathematical tool since $(H^2_-)^{p\times m}$ has a Hilbert space structure and is of interest in a stochastic context : the L^2 approximation problem is equivalent to minimize the mean square error between the output of a given system and the output of a model of fixed order when both systems have the same white noise input (see [10]).

In this paper, we generalize the results of [3] in the scalar case (single input single output) to the multivariable case. Recall that in the scalar case the problem can be reformulated as follows : eliminating the parameters in which the problem is linear, we are led to minimize a function Ψ_n defined on the set Δ_n of monic polynomials q of fixed degree n whose roots belong to $\mathbf{U}$. This set can be described by the coefficients of q and is open and bounded in $\mathbb{R}^n$; the function Ψ_n is smooth so that we can think of using a gradient algorithm. Moreover Ψ_n extends to a neighbourhood of Δ_n and has a nice behaviour : when we meet the boundary of Δ_n we are led to solve a problem of lower order, and the solution of a problem of order $k < n$ provides a boundary initial point for searching a minimum at order $k+1$. Thus, our procedure can continue through different orders until we find a local minimum at order n.

Transition from scalar to matrix-valued functions involves substantial new difficulties. Using the Douglas-Shapiro-Shields factorization (cf. [6]), we prove that we can restrict our search for an optimum to the set of normalized *inner functions* of degree n and size p, $\mathcal{I}_n^p(I)$, instead of Δ_n. This set has a structure of smooth manifold (see [1]), but which is not trivial any more. We give in this paper a self-contained and simplified proof of this result. It uses the theory of reproducing kernel Hilbert spaces (cf. [5]) and provides coordinates (the Schur parameters) to describe inner functions. A function of $\mathcal{I}_n^p(I)$ is computed by iterating a linear fractional transformation (the Schur transform) which changes an inner function into another one, the Mac-Millan degree being increased by one. One of the originality of this work is to give a useful representation of this transformation separating numerator and denominator. This representation allows us to make a complete study of the function Ψ_n on the boundary of $\mathcal{I}_n^p(I)$ and to prove that it presents the same behaviour than in the scalar case. The convergence to a local minimum has been generically proved.

Finally, an algorithm has been implemented, which is the first one to our knowledge which deals with matrix valued functions for the L^2-norm. It has already given good results in a great number of tests, and is under experimentation on a transfer function obtained from experimental data.

2 The function Ψ_n.

Let H^2 be the Hardy space of the open unit disk $\mathbf{U}$, while H^2_- consists of functions of the Hardy space for the complement of the closed unit disk vanishing at infinity, endowed with the usual L^2-norm. Indeed, each of these spaces can be identify with a closed subspace of L^2 of the unit circle $\mathbb{T}$, and we have the orthogonal decomposition

$$L^2(\mathbb{T}) = H^2 \oplus H^2_-.$$

Our transfer functions are assumed to be in $(H^2_-)^{p\times m}$, the space of matrices with entries in H^2_-. If $F = (f_{i,j})$ lies in $(H^2_-)^{p\times m}$, the square of its norm will be $\sum_{i,j} \|f_{i,j}\|_2^2$, where $\| \|_2$ is the L^2-norm. In this way, $(H^2_-)^{p\times m}$ becomes a Hilbert space whose norm and scalar product will be denoted like those of $L^2(\mathbb{T})$, i.e. $\| \|_2$ and $< , >$. If M is any matrix of $\mathbb{C}^{p\times m}$, we denote its transpose by ${}^tM$, its conjugate transpose by M^* and its trace by $\mathrm{Tr}M$. Let $\mathbb{C}^p$ and $\mathbb{C}^m$ be endowed with the Euclidean norm denoted by $\| \|$. We shall also denote by $\|M\|$ the operator norm of the multiplication by M from $\mathbb{C}^p$ to $\mathbb{C}^m$. If F and G lie in $(H^2_-)^{p\times m}$, their scalar product is

$$< F, G >= \frac{1}{2\pi}\mathrm{Tr}\int_0^{2\pi} F(e^{it})\overline{{}^tG(e^{it})}dt.$$

Taking into account the fact that $\bar{z} = z^{-1}$ on $\mathbb{T}$, this may be converted into a line integral :

$$< F, G >= \frac{1}{2i\pi}\mathrm{Tr}\int_{\mathbb{T}} F(z){}^t\bar{G}(1/z)\frac{dz}{z},$$

where $\bar{G}(z) = \overline{G(\bar{z})}$.

The problem of finding a strictly proper rational stable model for F can be stated as follows :
Given $F \in (H^2_-)^{p\times m}$, minimize the distance from F to the set of rational functions of degree exactly n, whose poles are in the unit disk.

This subset of $(H^2_-)^{p\times m}$ will be call Σ_n^-. To describe Σ_n^-, we shall use the **inner-unstable** or **Douglas-Shapiro-Shields factorization** (cf. [6] and [4]). Recall that an **inner** function Q of size p is a matrix of analytic functions in $\mathbf{U}$ which satisfies

$$Q(z)\ Q(z)^* \ =\ Q(z)^*\ Q(z)\ =\ I_p, \quad \forall z \in \mathbb{T}, \tag{1}$$

where I_p is the identity matrix of size p. We remark that any of the equalities in (1) implies the other. Then, we have

Proposition 1 *A transfer function T in $(H^2_-)^{p\times m}$ of degree n can be written*

$$T \ =\ Q^{-1}\ C, \tag{2}$$

where Q is a rational $p\times p$ inner function, C some rational matrix analytic in $\mathbf{U}$, and Q^{-1} has Mac-Millan degree n. The matrices Q and C may be chosen left coprime. With this condition, the factorization is unique up to a common left unitary factor.

To ensure uniqueness, we shall require that the inner factor satisfies the condition

$$Q(1) = I_p. \tag{3}$$

The set of rational inner functions of degree n will be denoted by $\mathcal{I}_n^p$, and by $\mathcal{I}_n^p(I)$ we denote the subset of functions satisfying the extra condition (3).

Now, we come to our approximation problem. Let $H = Q^{-1}C$ be a minimum of $\|F - R\|_2^2$, $R \in \Sigma_n^-$. Given Q, the matrix H belongs to the vector space $V_Q = (H_-^2)^{p\times m} \cap Q^{-1}(H^2)^{p\times m}$ of transfer functions whose left inner factor is equal to Q. Therefore, H must be the orthogonal projection of F onto V_Q. Thus C depends linearly on Q; it can be computed has the projection of QF onto $(H^2)^{p\times m}$, and will be denoted by $L(Q)$. Our approximation problem now consists in minimizing the function

$$\begin{array}{rcl} \Psi_n \ : \ \mathcal{I}_n^p(I) & \to & \mathbb{R} \\ Q & \to & \|F - Q^{-1}L(Q)\|_2^2 \end{array} \tag{4}$$

Let us explain how Ψ_n can be computed from Q and F. A rational representation of H can be given. To this end, we need the following lemma :

Lemma 1 *Let $Q \in \mathcal{I}_n^p$. Then, there exist a $p \times p$ polynomial matrix D of degree at most n, and a polynomial q of degree n whose roots lie in the unit disk $\mathbf{U}$, such that*

$$Q = D/\tilde{q}, \text{ and } Q^{-1} = \tilde{D}/q, \tag{5}$$

where, by definition, $\tilde{q}$ is the reciprocal polynomial of q,

$$\tilde{q}(z) = z^n\ \overline{q(1/\bar{z})} = z^n \bar{q}(1/z), \tag{6}$$

and

$$\tilde{D} \ = \ z^n\ {}^t\bar{D}(1/z). \tag{7}$$

Proof. Firstly, since Q^{-1} has degree n, its pole polynomial q also has degree n and qQ^{-1} is a polynomial matrix whose degree cannot exceed n (Q^{-1} is analytic outside the closed unit disk). In the scalar case ($p = 1$), Q^{-1} is a Blaschke product of the form $\epsilon\tilde{q}/q$, where ϵ is a complex number of modulus one. By analogy, we shall write $Q^{-1} \ = \ \tilde{D}/q$, where $\tilde{D}$ is the polynomial matrix deduced from some matrix D by (7). Now, since Q is inner and rational, it satisfies the relation $Q(z)^{-1} = {}^t\bar{Q}(1/z)$, so that $Q = D/\tilde{q}$. □

Now, we have

Proposition 2 *The function Ψ_n can be computed at Q by*

$$\Psi_n(Q) \ = \ \|F\|_2^2 - < F, \frac{\tilde{P}}{q} >, \tag{8}$$

where $\tilde{P} = z^{n-1}\ {}^t\bar{P}(1/z)$ and P is a polynomial matrix of degree at most $n-1$, given by

$$P \ = \ \frac{RD}{q}, \tag{9}$$

where D and q are associated with Q as in (5), and R is the remainder in the Weierstrass division in $(H^2)^{p\times m}$ of $G(z) = {}^t\bar{F}(1/z)/z$ by q :

$$G\tilde{D} \ = \ V\ q \ + \ R. \tag{10}$$

Proof. Since H and $F - H$ are orthogonal,

$$\Psi_n(Q) = \|F\|_2^2 - < F, Q^{-1}L(Q) > .$$

The orthogonal projection $L(Q)$ of QF onto $(H^2)^{p\times m}$ is easily computed from division (10). Indeed, applying to equation (10) the transformation $M \to z^{n-1}\ {}^t\bar{M}(1/z)$, yields $L(Q) \ = \ \frac{\tilde{R}}{\tilde{q}}$, where $\tilde{R} \ = \ z^{n-1}\ {}^t\bar{R}(1/z)$. Now, multiplying (10) by D on the right shows that q divides RD, so that H is in $(H^2_-)^{p\times m}$. Thus (9) defines a polynomial matrix of degree at most $n-1$, and $H \ = \ \frac{\tilde{P}}{q}$. □

Now, we must study the structure of $\mathcal{I}_n^p(I)$. This will be done by a now classical matrix version of the Schur algorithm. The results presented in the next two sections are elaborate on [1].

3 Reproducing kernel Hilbert spaces.

A complex Hilbert space H of $\mathbb{C}^p$-valued functions defined on some Ω open in $\mathbb{C}$ is called a **reproducing kernel Hilbert space** if there exists a $\mathbb{C}^{p\times p}$-valued function $K(z,w)$ defined on $\Omega \times \Omega$ such that

(i) $\forall c \in \mathbb{C}^p$, $\forall w \in \Omega$, the function

$$\begin{array}{rcl} K(.,w)c \ : \ \Omega & \to & \mathbb{C}^p \\ z & \to & K(z,w)c \end{array}$$

belongs to H.

(ii) $\forall f \in H$, $\forall w \in \Omega$, $\forall c \in \mathbb{C}^p$, we have

$$< f, K(.,w)c > = \ c^* f(w), \tag{11}$$

where $< .,. >$ denotes the scalar product in H, and $c^* \ = \ {}^t\bar{c}$.

The function K is called the **reproducing kernel**, and is unique.

The Hardy space $(H^2)^p$ is clearly a reproducing kernel Hilbert space whose kernel is

$$\frac{I_p}{1-\bar{w}z},$$

and property (ii) is just the Cauchy formula. Finite dimensional Hilbert spaces of $\mathbb{C}^p$-valued functions are also reproducing kernel Hilbert spaces. Let $(f_1, f_2, \ldots, f_N)$ be some base of a finite dimensional Hilbert space. Then its reproducing kernel is easily computed to be

$$K(z,w) = (f_1(z), f_2(z), \ldots, f_N(z))P^{-1}(f_1(w), f_2(w), \ldots, f_N(w))^*,$$

where $P = (P_{ij})$ is the Gram matrix with entries $P_{ij} \; = < f_j, f_i >$.

In the sequel we shall be concerned with two reproducing kernel Hilbert spaces associated with inner matrices. The first one, denoted by $H(Q)$ where $Q \in \mathcal{I}_n^p$, is the orthogonal complement of $Q(H^2)^p$ in $(H^2)^p$:

$$(H^2)^p \;=\; Q(H^2)^p \;\oplus\; H(Q). \tag{12}$$

We have $H(Q) = Q(H^2_-)^p \cap (H^2)^p$, so that we obtain the projection of some f onto $H(Q)$ by projecting $Q^{-1}f$ onto $(H^2_-)^p$ and multiplying by Q. We know that $H(Q)$ is an Hilbert subspace of $(H^2)^p$ of dimension n, the degree of Q. We have, $\forall f \in H(Q)$,

$$< \frac{I_p c}{1-\bar{w}z}, f > \;=\; c^* f(w),$$

but of course, the function $I_p c/1-\bar{w}z$ does not belong to $H(Q)$. However, when we take its projection onto $H(Q)$, since

$$\Pi^+(Q(z)^{-1}\frac{I_p c}{1-\bar{w}z}) \;=\; \frac{Q(w)^* c}{1-\bar{w}z},$$

where Π^+ denotes the projection onto $(H^2)^p$, we obtain

$$\frac{(I_p - Q(z)Q(w)^*)c}{1-\bar{w}z},$$

and this function satisfies conditions **(i)** and **(ii)** of a reproducing kernel. Thus $H(Q)$ is the reproducing kernel Hilbert space with reproducing kernel

$$K_Q(z,w) \;=\; \frac{(I_p - Q(z)Q(w)^*)c}{1-\bar{w}z}. \tag{13}$$

Let us introduce the second reproducing kernel Hilbert space that we need, closely linked with the previous one. Let

$$J \;=\; \begin{pmatrix} I_p & 0 \\ 0 & -I_p \end{pmatrix}.$$

Consider the space $(L^2(\mathbb{T}))^{2p}$ endowed with the sesquilinear hermitian form, $\forall\ f,\ g$ in $(L^2(\mathbb{T}))^{2p}$,

$$< f, g >_J = < f, Jg > .$$

This form is not positive definite, so that $(L^2(\mathbb{T}))^{2p}$ is not an Hilbert space, but it is non-degenerate. We shall consider $(H^2)^{2p}$ when endowed with this form. In that space, inner functions are to be replaced by J-inner functions. A rational function $\Theta(z)$ analytic in $\mathbf{U}$ is called **J-inner** if it satisfies the condition

$$\Theta(z)^* J \Theta(z) \ = \ \Theta(z) J \Theta(z)^* \ = \ J, \quad \forall z \in \mathbb{T}. \tag{14}$$

Since the form $< .,. >_J$ is non-degenerate, the space $\Theta(H^2)^{2p}$ has an orthogonal complement in $(H^2)^{2p}$, which we call $H(\Theta)$, that is

$$(H^2)^{2p} \ = \ \Theta(H^2)^{2p} \ \oplus_J \ H(\Theta).$$

Moreover the intersection $\Theta(H^2)^{2p} \cap H(\Theta)$ is zero, and $H(\Theta)$ satisfies

$$H(\Theta) \ = \ \Theta J(H^2_-)^{2p} \cap (H^2)^{2p}.$$

Restricted to $H(\Theta)$, the form $< .,. >_J$ is positive definite, so that $H(\Theta)$ is an Hilbert space. As in the case of $H(Q)$, we can see that it is a reproducing kernel Hilbert space with reproducing kernel

$$K_\Theta(z,w) \ = \ \frac{(J - \Theta(z) J \Theta(w)^*)}{1 - \bar{w}z}. \tag{15}$$

and that the dimension of $H(\Theta)$ agrees with the Mac-Millan degree of Θ.

To conclude this section, we shall give an important property of reproducing kernel : reproducing kernels are **positive functions** in the following sense :
$\forall r$, integer, $\forall w_1, w_2, \ldots, w_r \in \Omega$, $\forall c_1, c_2, \ldots, c_r \in \mathbb{C}^p$,

$$\sum_{i,j=1}^{r} c_j^* K(w_j, w_i) c_i \geq 0. \tag{16}$$

Moreover, there is a one-to-one correspondence between positive functions and reproducing kernel Hilbert spaces (see [12]). In particular, the kernels (13) and (15) are positive, so that, for all z in $\mathbf{U}$, the matrices $I_p - Q(z)Q(z)^*$ and $J - \Theta(z)J\Theta(z)^*$ are positive semidefinite, in other words:

$$Q(z)Q(z)^* \leq I_p, \quad \forall z \in \mathbf{U} \tag{17}$$

$$\Theta(z)J\Theta(z)^* \leq J, \quad \forall z \in \mathbf{U} \tag{18}$$

These properties are known as contractivity of inner functions and J-contractivity of J-inner functions, and, they can be deduced from the analyticity in $\mathbf{U}$ (and the maximum principle) and properties (1) and (14) on the unit circle.

In the sequel, we will also use the following property : a reproducing kernel Hilbert space is the completion of the linear space generated with the functions $z \to K(z,w)c$, $w \in \Omega$, $c \in \mathbb{C}^p$.

Note that J-inner functions can be defined in a larger context, and may have poles in $\mathbf{U}$, but in this case condition (18) must be included in the definition. In this paper, all J-inner functions are analytic in $\mathbf{U}$.

In the following section, we shall see how J-inner functions and $H(\Theta)$ spaces may arise in the construction of inner functions and, in particular, the one dimensional spaces. The following is easily verified :

Proposition 3 *Let Θ be some J-inner function of degree 1. Then, $H(\Theta)$ is generated by a single element of the form*

$$\frac{\begin{pmatrix} u \\ v \end{pmatrix}}{1-\bar{w}z},$$

where $w \in \mathbf{U}$, $u \in \mathbb{C}^p$, $\|u\| = 1$, $v \in \mathbb{C}^p$, $\|v\| \leq 1$ (so that the inner product $<\ ,\ >_J$ is positive definite), and Θ is given, up to some constant J-inner right factor, by

$$\Theta(z) \;=\; I_{2p} - \frac{1}{1-\|v\|^2}\frac{(1-z)(1-\bar{w}w)}{(1-w)(1-\bar{w}z)}\begin{pmatrix} u \\ v \end{pmatrix}\begin{pmatrix} u^* & v^* \end{pmatrix} J. \tag{19}$$

When necessary, we shall indicate the dependance in w,u, and v by the notation $\Theta(w,u,v)$.

4 The Schur transform.

Let A be some inner function of size p and Θ a $2p \times 2p$ matrix, rational and analytic in $\mathbf{U}$:

$$\Theta \;=\; \begin{pmatrix} \Theta_1 & \Theta_2 \\ \Theta_3 & \Theta_4 \end{pmatrix}. \tag{20}$$

Consider the linear fractional transformation

$$T_\Theta(A) \;=\; (\Theta_1 A + \Theta_2)\,(\Theta_3 A + \Theta_4)^{-1}. \tag{21}$$

If det $(\Theta_3 A + \Theta_4)$ does not vanish identically, this defines a new rational matrix of size p, $B \;=\; T_\Theta(A)$. The question is : is B also inner? A partial answer is : if Θ is assumed to be J-inner, it will be, as we now show : so suppose that Θ is J-inner.

At first, let us show that $(\Theta_3 A + \Theta_4)$ is invertible at every point of $\mathbf{U}$. Indeed, condition (18) implies

$$\Theta_4 \Theta_4^* \;\geq\; I_p + \Theta_3 \Theta_3^*, \quad \text{in } \mathbf{U},$$

so that $\Theta_4 \Theta_4^*$ is positive definite, and Θ_4 is invertible at any point of $\mathbf{U}$. Now, we have

$$I_p \;\geq\; \Theta_4^{-1}(\Theta_4^{-1})^* + (\Theta_4^{-1}\Theta_3)(\Theta_4^{-1}\Theta_3)^* \quad \text{in } \mathbf{U},$$

so that

$$\|\Theta_4(z)^{-1}\Theta_3(z)\| \;<\; 1, \quad \forall z \in \mathbf{U}.$$

The matrix A, being inner, is contractive in $\mathbf{U}$ and

$$\|A(z)\| \;\leq\; 1, \quad \forall z \in \mathbf{U},$$

so that

$$\|\Theta_4(z)^{-1}\Theta_3(z)A(z)\| \;<\; 1, \quad \forall z \in \mathbf{U}.$$

Finally, $(\Theta_3 A + \Theta_4) \;=\; \Theta_4\ (I_p + \Theta_4^{-1}\Theta_3 A)$ is invertible at any point of $\mathbf{U}$, and thus B is analytic in $\mathbf{U}$.

Now, condition (1) for B becomes

$$B^*B - I_p \;=\; \begin{pmatrix} B^* & I_p \end{pmatrix} J \begin{pmatrix} B \\ I_p \end{pmatrix} \;=\; 0, \quad \text{on } \mathbb{T}.$$

But since

$$\begin{pmatrix} B \\ I_p \end{pmatrix} \;=\; \Theta \begin{pmatrix} A \\ I_p \end{pmatrix} (\Theta_3 A + \Theta_4)^{-1}, \tag{22}$$

$$B^*B - I_p \;=\; ((\Theta_3 A + \Theta_4)^{-1})^* \begin{pmatrix} A^* & I_p \end{pmatrix} \Theta^* J \Theta \begin{pmatrix} A \\ I_p \end{pmatrix} (\Theta_3 A + \Theta_4)^{-1},$$

and since condition (1) is satisfied for A, it will be satisfied for B as well.

Note that, if $\Theta(1)$ is the identity matrix, then $A(1) = I_p$ implies $B(1) = I_p$, and if A and Θ have real coefficients, B also has real coefficients.

In the next section, we shall work with J-inner matrices $\Theta(w, u, v)$ of the form (19). In this case, we have

Lemma 2 *The matrix $B \;=\; T_{\Theta(w,u,v)}(A)$, satisfies the interpolation condition*

$$B(w)^* u \;=\; v. \tag{23}$$

Proof. Indeed, it can be verified that $\Theta(w,u,v)$ satisfies the equation

$$\begin{pmatrix} u^* & -v^* \end{pmatrix} \Theta(w) \;=\; 0.$$

Thus

$$\begin{pmatrix} u^* & -v^* \end{pmatrix} \Theta(w) \begin{pmatrix} A(w) \\ I_p \end{pmatrix} \;=\; 0,$$

and by (22), this implies our interpolation condition. □

Now, the question is the converse : let B be some rational inner matrix, and Θ J-inner analytic in $\mathbf{U}$. Can we write B in the form $B \;=\; T_\Theta(A)$ for some inner matrix A? The theory of reproducing kernel Hilbert spaces gives (cf. [5]) a beautiful answer to this question.

First, remark that if $B \;=\; T_\Theta(A)$ then A is the rational function given by

$$A \;=\; (\Theta_1 - B\Theta_3)^{-1}(B\Theta_4 - \Theta_2), \tag{24}$$

unless $\det(\Theta_1 - B\Theta_3)$ vanishes identically. But this may not happen, since conditon (18) for Θ implies

$$\Theta_1^*\Theta_1 - \Theta_3^*\Theta_3 \;=\; I_p \quad \text{on } \mathbb{T}.$$

So, $\Theta_1 - B\Theta_3$ is invertible at any point of $\mathbb{T}$. But A may fail to be invertible at some point of $\mathbf{U}$, so that A may not be analytic in $\mathbf{U}$. To ensure analyticity, we must make an additional assumption.

Consider the map

$$\begin{array}{rcl} \tau \;:\; H(\Theta) & \rightarrow & H_2^p \\ f \;=\; \begin{pmatrix} f_1 \\ f_2 \end{pmatrix} & \rightarrow & f_1 - Bf_2. \end{array}$$

It is a bounded operator. If Θ is given by (19), we have

Lemma 3 *The conditions*
(i) $\tau(H(\Theta(w,u,v))) \subset H(B)$,
(ii) $B(w)^*u = v$,
are equivalent. If they are satisfied, τ defines an isometry from $H(\Theta)$ to $H(B)$.

Proof. Indeed, $f_1 - Bf_2$ belongs to $H(B)$ if and only if $f_2 \;=\; \Pi^+(B^{-1}f_1)$, and

$$\Pi^+(B^{-1}\frac{u}{1-\bar{w}z}) \;=\; \frac{B(w)^*u}{1-\bar{w}z},$$

which proves the equivalence. In this case,

$$< f, f >_J \;=\; \|f_1\|_2^2 - \|f_2\|_2^2 \;=\; \|f_1\|_2^2 - \|Bf_2\|_2^2 \;=\; \|f_1 - Bf_2\|_2^2,$$

and τ is an isometry from $H(\Theta)$ endowed with the scalar product of $(H^2_J)^{2p}$ to $H(B)$ endowed with that of $(H^2)^p$. $\square$

Theorem 1 *Let B be a rational inner function, and Θ J-inner analytic in $\mathbf{U}$. There exists an inner function A such that $B = T_\Theta(A)$, if and only if, $\tau(H(\Theta)) \subset H(B)$. Moreover,* $\deg B = \deg\Theta + \deg A$.

Proof. Let τ^* be the adjoint of τ. Using properties of the reproducing kernel , we obtain

$$\tau^*(K_B(.,w)c) = K_\Theta(.,w)\begin{pmatrix} I_p \\ -B(w)^* \end{pmatrix} c, \tag{25}$$

which gives τ^* on $H(B)$. Using (25), it can be verified that

$$[(I_p - \tau\tau^*)K_B(.,w)c](z) = \frac{[(\Theta_1 - B\Theta_3)(z)(\Theta_1 - B\Theta_3)(w)^* - (\Theta_2 - B\Theta_4)(z)(\Theta_2 - B\Theta_4)(w)^*]c}{1 - \bar{w}z}. \tag{26}$$

Now, assume that

$$\tau(H(\Theta)) \subset H(B).$$

In this case, τ shall also denote the map

$$\tau : H(\Theta) \to H(B).$$

This is clearly an isometry, and thus it induces an orthogonal decomposition

$$H(B) = \tau(H(\Theta)) \oplus (I_p - \tau\tau^*)H(B). \tag{27}$$

Since τ is an isometry, the operator $I_p - \tau\tau^*$ is positive. Thus, applied to w, equation (26) implies that the matrix

$$(\Theta_1 - B\Theta_3)(\Theta_1 - B\Theta_3)^* - (\Theta_2 - B\Theta_4)(\Theta_2 - B\Theta_4)^*$$

is positive semi-definite at any point of $\mathbf{U}$. Thus the matrix A given by (24) satisfies

$$I_p - A(w)A(w)^* \geq 0,$$

at any point w of its domain of definition, that is almost everywhere on $\mathbf{U}$. But, this implies that A is bounded almost everywhere and thus defined and analytic in $\mathbf{U}$.

Let us now prove the converse. At first, suppose that Θ is given by (19). If there exists A inner such that $B = T_\Theta(A)$, then by Lemma 2, $B(w)^*u = v$, and by Lemma 3, $\tau(H(\Theta)) \subset H(B)$. Moreover, equality (26) asserts that

$$(I_p - \tau\tau^*)K_B(.,w)c = (\Theta_1 - B\Theta_3)K_A(.,w)(\Theta_2 - B\Theta_4)(w)^*c. \tag{28}$$

On the other hand, the functions $K_B(.,w)c$, $w \in \mathbf{U}$, $c \in \mathbb{C}^p$ are dense in $H(B)$, while the functions $K_A(.,w)(\Theta_2 - B\Theta_4)(w)^*c$ are dense in $H(A)$ (since $(\Theta_2 - B\Theta_4)(w)^*c$, $c \in \mathbb{C}^p$, describes all of $\mathbb{C}^p$ except for a finite number of w's). Thus we have

$$(I_p - \tau\tau^*)H(B) \;=\; (\Theta_1 - B\Theta_3)H(A), \tag{29}$$

and these spaces have dimension $\deg A$. Finally, the orthogonal decomposition (27) asserts that

$$\deg B = \deg A + 1,$$

and the theorem is proved in this case.

Now let Θ be any J-inner function, and suppose that $B = T_\Theta(A)$ for some inner function A. First, we shall prove that

$$\deg B = \deg A + \deg\Theta.$$

If Θ has degree one, this is obviously true. Let $\deg\Theta = d$. By the Potapov decomposition (cf. [11]),

$$\Theta = \Theta_1\Theta_2 \ldots \Theta_d,$$

where the Θ_j, $j = 1, \ldots, d$ are J-inner functions of degree one. Since

$$T_\Theta \;=\; T_{\Theta_1} \circ T_{\Theta_2} \ldots . \circ T_{\Theta_d},$$

and the degree increases by one for any application of T_{Θ_j}, the equality on the Mac-Millan degrees is satisfied.

Now we have

$$H(B) \;=\; \tau\tau^* H(B) + (I_p - \tau\tau^*)H(B),$$

and thus

$$H(B) \;\subset\; \tau H(\Theta) + (I_p - \tau\tau^*)H(B). \tag{30}$$

We still have

$$(I_p - \tau\tau^*)H(B) \;=\; (\Theta_1 - B\Theta_3)H(A),$$

so that the dimension of $(I_p - \tau\tau^*)H(B)$ equals $\deg A$, while that of $\tau H(\Theta)$ does not exceed $\deg\Theta$. But the dimension of $H(B)$ equals $\deg B = \deg A + \deg\Theta$. Therefore, in (30) the sum is direct and the inclusion an equality :

$$H(B) \;=\; \tau H(\Theta) \oplus (I_p - \tau\tau^*)H(B).$$

This proves that $\tau(H(\Theta)) \subset H(B)$. $\square$

5 An atlas for $\mathcal{I}_n^p(I)$.

In [1] two atlases of $\mathcal{I}_n^p(I)$ are constructed, originated on transformation (21). We choose to work with the second one, since the range of the charts is then the product of n copies of the unit ball, which is more convenient. These charts are obtained by using a tangential Schur algorithm that we shall now describe.

If Θ is of the form (19), then $T_\Theta(A)$ and A have the same value at $z = 1$. We can construct from the identity matrix I_p, using n linear fractional transformations of this type, an inner matrix of degree n satisfying (3), and thus which belongs to $\mathcal{I}_n^p(I)$. Conversely, any matrix in $\mathcal{I}_n^p(I)$ can be constructed in this way. To show this, we must first invert the linear fractional transformation T_Θ. Theorem 1 asserts that you can, if and only if the interpolation condition (23) is satisfied. To this end, we must find w and u as in

Lemma 4 *Let B be an inner function and $w \in \mathbf{U}$. Then,*

$$\exists u \in \mathbb{C}^p, \quad \|B(w)^* u\| < 1.$$

Proof. Suppose that

$$\forall u \in \mathbb{C}^p, \quad \|B(w)^* u\| = 1.$$

Then, for all $u \in \mathbb{C}^p$,

$$\|K_B(.,w)u\|_2^2 \;=\; u^* K_B(w,w)u \;=\; \frac{1 - \|B(w)^* u\|}{1 - \bar{w}w} \;=\; 0,$$

and

$$K_B(.,w)u \;=\; 0.$$

So, $K_B(.,w) \;=\; 0$ and the matrix B must be constant. But this contradicts the fact that B has Mac-Millan degree n. □

Let $B \in \mathcal{I}_{k+1}^p(I)$, w, u as in Lemma 4, and put $y = B(w)^* u$. Then, by Theorem 1, there exists $A \in \mathcal{I}_k^p(I)$ such that $B \;=\; T_{\Theta(w,u,y)}(A)$.

The tangential Schur algorithm consists in repeating this construction. Let $Q \in \mathcal{I}_n^p(I)$, and $w_1, \ldots, w_n,\ w_k \in \mathbf{U}$. Then, for $k = 1$ to n there exist unit vectors u_k in $\mathbb{C}^p$ such that,

$$\|Q^{(k)}(w_k)^* u_k\| < 1,$$

where the matrix $Q^{(k-1)}$ is defined from $Q^{(k)}$, $Q^{(n)}$ being equal to Q, by

$$Q^{(k)} \;=\; T_{\Theta^{(k)}}(Q^{(k-1)}) \tag{31}$$

with

$$\Theta^{(k)} = \Theta(w_k, u_k, y_k), \tag{32}$$

and

$$y_k \;=\; [Q^{(k)}(w_k)]^* u_k. \tag{33}$$

We have $Q^{(1)} = T_{\Theta^{(1)}}(I)$, the matrix $Q^{(k)}$ lies in $\mathcal{I}_k^p(I)$, and finally

$$Q \;=\; T_{\Theta^{(n)}}(T_{\Theta^{(n-1)}} \dots T_{\Theta^{(1)}}(I_p)) \dots).$$

Let $\mathbf{w} = (w_1, w_2, \dots, w_n)$, and $\mathbf{u} = (u_1, u_2, \dots, u_n)$. Define the subset $\mathcal{V}_{(\mathbf{w},\mathbf{u})}$ of $\mathcal{I}_n^p(I)$ by

$$\mathcal{V}_{(\mathbf{w},\mathbf{u})} \;=\; \{Q \in \mathcal{I}_n^p(I) \;/\; \|Q^{(k)}(w_k)^* u_k\| < 1\},$$

and the function $\phi_{(\mathbf{w},\mathbf{u})}$ by

$$\begin{array}{rcl} \phi_{(\mathbf{w},\mathbf{u})} \;:\; \mathcal{V}_{(\mathbf{w},\mathbf{u})} & \rightarrow & B(\mathbb{C}^p)^n \\ Q & \rightarrow & (y_1, y_2, \dots, y_n) \end{array},$$

where the $Q^{(k)}$'s and the y_k's are defined recursively by (31), (32) and (33).

It follows from Lemma 4 that the collection of sets $\mathcal{V}_{(\mathbf{w},\mathbf{u})}$ covers $\mathcal{I}_n^p(I)$. Let $\mathcal{V}_i$, i ranging in some indexing set, be any refinement of this covering, and ϕ_i the map defined on $\mathcal{V}_i$. We then have

Theorem 2 *The family $(\mathcal{V}_i, \phi_i)$ defines a C^∞ atlas on $\mathcal{I}_n^p(I)$ which is compatible with its natural structure of embedded submanifold of $(H^2)^{p\times p}$.*

Proof. We have just shown that the sets $\mathcal{V}_i$ cover $\mathcal{I}_n^p(I)$. On the other hand, the map

$$\begin{array}{rcl} \mathcal{S}_{(w,u)} \;:\; \mathcal{I}_k^p \times B(\mathbb{C}^p) & \rightarrow & \mathrm{Im}\, \mathcal{S}_{(w,u)} \\ (A, y) & \rightarrow & B = T_{\Theta(w,u,y)}(A), \end{array} \tag{34}$$

is an homeomorphism. Indeed, B is a rational function of the coefficients of A and y, and in $(H^2)^{p\times p}$, if a sequence of normalized rational functions of bounded degree converges to some rational function, then the coefficients also converge. Since the matrix B is inner and thus bounded in the unit disk, $\|B(z)\| \leq 1$, $\forall z \in \mathbf{U}$, the Lebesgue theorem implies that $\mathcal{S}_{(w,u)}$ is continuous. In the same way, A depends continuously on B and moreover, the mapping $B \rightarrow B(w)^* u$ is continuous, so that $\mathcal{S}_{(w,u)}$ is an homeomorphism. So, it follows that the ϕ_i are homeomorphisms.

Furthermore, the map

$$\phi_i \circ \phi_j \;:\; B(\mathbb{C}^p)^n \rightarrow B(\mathbb{C}^p)^n,$$

is a rational function, and thus is C^∞. □

Remark : A covering of $\mathcal{I}_n^p(I)$ that we shall use in practice is obtained in the following way :
1) $\mathbf{w} = (w, \ldots, w)$, where w is fixed in $\mathbf{U}$; in particular, $w = 0$ is a nice choice.
2) the components of the $\mathbf{u}$'s are chosen among n independent unit vectors of $\mathbb{C}^p$, for instance $e_1, \ldots, e_n$, where $e_k = (\delta_k^1, \ldots, \delta_k^n)$.

The y_k's defined by (31) and (33) are called the **Schur parameters**.

In practice, transfer functions have real Fourier coefficients. In this case the inner functions involved in the inner-unstable factorizations also have real coefficients, so that we may need a parametrization of the set $\mathcal{RI}_n^p(I)$ of functions of $\mathcal{I}_n^p(I)$ that have real coefficients.

This case can be viewed as a particular case of the complex one. Indeed, in Lemma 4 u can be chosen real, and if A and Θ have real coefficients, $T_\Theta(A)$ also has real coefficients. We can eventually construct an atlas for $\mathcal{RI}_n^p(I)$, for which the w_i's lie in $]-1,1[$, the u_i's and the y_i's have real components. The range of the charts is thus the product of n copies of the unit ball of $\mathbb{R}^p$.

6 Study of a Schur iteration.

We shall call Schur transform the map $\mathcal{S}_{(w,u)}$ defined in (34). It will be useful in the sequel to have an explicit representation for $\mathcal{S}_{(w,u)}(A,y)$. More generally this representation can be computed for any $A \in \mathcal{I}_k^p$. Using (19) yields

$$\mathcal{S}_{(w,u)}(A,y) \;=\; \left(I_p + (1-\beta_w)\frac{u(y^*-u^*A)}{1-y^*y}\right)\;\left(I_p + (1-\beta_w)\frac{y(y^*-u^*A)}{1-y^*y}\right)^{-1},$$

with

$$\beta_w = \frac{b_w}{\tilde{b}_w},$$

where

$$b_w(z) = (z-w)(1-\bar{w}).$$

Consider the mapping of $\mathbb{C}^p$ into $\mathbb{C}^p$

$$T \;:\; v \to v + (1-\beta_w)\frac{(y^*-u^*A)v}{1-y^*y}y.$$

If v is orthogonal to $u - Ay$ then $T(v) = v$, and $T(y) = \left(1 + (1-\beta_w)\frac{(y^*-u^*A)y}{1-y^*y}\right) y$. So, we have two T-invariant complementary subspaces of $\mathbb{C}^p$. From these considerations, we can compute the inverse mapping to be

$$T^{-1}:\; v \to v - (1-\beta_w)\frac{(y^* - u^*A)v}{1 - u^*Ay - \beta_w(y^*y - u^*Ay)}y.$$

We finally have

$$\mathcal{S}_{(w,u)}(A,y) \;=\; A + \frac{1-\beta_w}{1-u^*Ay-\beta_w(y^*y-u^*Ay)}(u-Ay)(y^*-u^*A). \tag{35}$$

Write A in the form $A = D_A/\tilde{q}_A$ (cf. Lemma 1), we get for B the same form. Indeed, $B = D_B/\tilde{q}_B$ where

$$\begin{aligned} D_B &= (\tilde{b}_w - y^*y\, b_w)D_A \\ &\quad -(\tilde{b}_w - b_w)\left(u\; u^*D_A + D_Ay\; y^* - \tilde{q}_A\; u^*y + \frac{u^*D_Ay\; D_A - D_Ay\; u^*D_A}{\tilde{q}_A}\right) \quad (36) \\ \tilde{q}_B &= (\tilde{b}_w - y^*y\, b_w)\tilde{q}_A - (\tilde{b}_w - b_w)\, u^*D_Ay. \quad (37) \end{aligned}$$

Remark : in equation (36), $\tilde{q}_A$ does indeed divide $u^*D_Ay\; D_A - D_Ay\; u^*D_A$. To see this, put

$$u = {}^t(u_1,\ldots,u_p); \quad y = {}^t(y_1,\ldots,y_p), \quad D_A = (d_{ij}).$$

A straight forward computation shows that

$$(u^*D_Ay\; D_A - D_Ay\; u^*D_A)_{ij} = \sum_{l,m}(d_{lm}d_{ij} - d_{im}d_{lj})u_l y_m,$$

where $d_{lm}d_{ij} - d_{im}d_{lj}$ is a minor of order 2 of D_A, and since A is inner, $\tilde{q}_A$ divides all the minors of order 2 of D_A.

Formulas (35), (36) and (37) are given for Schur parameters of norm strictly less than 1. However, it is of importance for later usage to note that these formulas remain defined also if some Schur parameter has a unit norm, while $u^*Ay \neq 1$, and even if this norm become greater than 1. Of course, in this case, they do not have the same interpretation. An important result is that $\mathcal{S}_{(w,u)}$ extends continuously in the neighbourhood of a Schur parameter with unit norm such that $u^*Ay \neq 1$.

Proposition 4 *Let (y_i) be a sequence in $B(\mathbb{C}^p)$ converging to some y, such that $\|y\| = 1$, and (A_i) a sequence in $\mathcal{I}_k^p$ converging to some $A \in \mathcal{I}_k^p$. Let $B_i = S_{(w,u)}(A_i, y_i)$.*
(i) if $u^*Ay \neq 1$, then (B_i) converges in $(H^2)^{p\times p}$ to the inner function of degree at most k :

$$B \;=\; \mathcal{S}_{(w,u)}(A,y) \;=\; A + \frac{(u-Ay)(y^*-u^*A)}{(1-u^*Ay)}. \tag{38}$$

(ii) if $u^*Ay = 1$ identically, then (B_i) has no limit.

Proof. Let us suppose that the sequence (A_i) is constant; the result for any sequence will then follow from the continuity of the map $A \to S_{(w,u)}(A, y)$. Formula (35) yields

$$B_i \;=\; A + \frac{1-\beta_w}{(1-\beta_w)(1-u^*Ay_i)+\beta_w(1-y_i^*y_i)}(u - Ay_i)(y_i^* - u^*A).$$

(i) if $u^*Ay \neq 1$, then (B_i) converges to (38) at every point of $\mathbf{U}$. The matrix B is inner and, from the Lebesgue theorem, the sequence (B_i) converges to B in the L^2-norm. The denominators q_{B_i} of the $(B_i)^{-1}$'s converge to

$$q_B = (b_w - \tilde{b}_w)(q - y^*\tilde{D}_A u),$$

while the numerators $\tilde{D}_{B_i}$ converge to

$$\tilde{D}_B \;=\; (b_w - \tilde{b}_w)\left(\tilde{D}_A - \tilde{D}_A u\ u^* - y\ y^*\tilde{D}_A + q_A\ yu^* - \frac{y^*\tilde{D}_A u\ \tilde{D}_A - \tilde{D}_A u\ y^*\tilde{D}_A}{q_A}\right).$$

We first notice that q_B has degree $k+1$. Indeed, this will be so if $\tilde{q}_B(0) \neq 0$. But,

$$\tilde{q}_B(0) = (1-|w|^2)\tilde{q}_A(0)(1 - u^*A(0)y),$$

is not zero, because $\tilde{q}_A(0) \neq 0$ and the maximum modulus principle applied to the analytic function y^*Au implies that $1 - u^*A(0)y \neq 0$. Next, q_B has its roots in $\overline{\mathbf{U}}$. In particular, 1 is the unique root of $b_w - \tilde{b}_w$ and thus of q_B. But the factor $(z-1)$ simplifies in B, since it also appears in $\tilde{D}_B$. So, the degree of B does not exceed that of A. Since B is inner, this simplification occurs for any root of modulus one of q_B. Now, ξ is a root of modulus one of q_B if and only if it satisfies $1 = y^*A(\xi)^{-1}u = y^*A(\xi)^*u$, that is $y = A(\xi)^*u$. Finally, we have

$$\deg B = \deg A - \mathrm{card}\{\xi \in \mathbb{T}, y = A(\xi)^*u\}.$$

Moreover,

$$B(1) = \left(I_p + \frac{(u - A(1)y)(y^*A(1)^* - u^*)}{(1 - u^*A(1)y)}\right)A(1)$$

does not equal $A(1)$.

(ii) if $u^*Ay = 1$ identically, then u^*A is a constant vector equal to y^*. Thus,

$$q_{B_i} = ((b_w - \tilde{b}_w)(1 - y_i^*y) + \tilde{b}_w(1 - y_i^*y))q_A,$$

so that $q_{B_i} \to 0$ as $y_i \to y$ and the limit of B_i does not exist. Let $\lambda \in]0,1[$, and S_λ be the hypersurface in $\mathbb{C}^p$ defined by the equation

$$x \in \mathcal{S}_\lambda \Leftrightarrow 1 - x^*x = 2\lambda\|1 - x^*y\|.$$

Suppose that y_i converges to y with the y_i's in S_λ :

$$q_{B_i} = (b_w + (2\lambda - 1)\tilde{b}_w)(1 - y_i^*y)q_A.$$

Then

$$\det B_i = \frac{(b_w + (2\lambda - 1)\tilde{b}_w)}{(\tilde{b}_w + (2\lambda - 1)b_w)} \det A,$$

and so B_i converges to some inner function of degree n depending on λ. □

Remark : in practice, it is not easy to check condition (ii). However, it follows from the proof that the three conditions
1) $u^*Ay = 1$ identically,
2) $u^*A(0)y = 1$,
3) $\tilde{q}_B(0) = 0$,
are equivalent. As we shall see, the 3rd one will be particularly convenient.

7 A generic algorithm to find a local minimum.

The closure of $\mathcal{I}_n^p(I)$ in $(H^2)^{p\times p}$, $\overline{\mathcal{I}_n^p(I)}$, is a compact set, so that we can think of using a gradient algorithm to find a local minimum of Ψ_n defined in (4). The elements of $\mathcal{I}_n^p(I)$ will be represented by the Schur parameters described in the previous section. Thus, we shall work with the restrictions of Ψ_n to the ranges of the charts. Let $(\mathcal{V}_n, \phi_n)$ be some chart of $\mathcal{I}_n^p(I)$, and

$$\begin{array}{llllll} \Psi_{|\mathcal{V}_n}: & B(\mathbb{C}^p)^n & \rightarrow & \mathcal{V}_n \subset \mathcal{I}_n^p(I) & \rightarrow & \mathbb{R} \\ & \mathbf{y} = (y_1, \dots y_n) & \rightarrow & \phi_n^{-1}(\mathbf{y}) = Q & \rightarrow & \Psi_n(Q) = \|F - Q^{-1}L(Q)\|_2^2. \end{array}$$

Let $Q^{-1} = \tilde{D}/q$, where $\tilde{D} = \tilde{D}^{(n)}$ and $q = q^{(n)}$ are computed recursively from the Schur formulas (36) and (37) :

$$\begin{aligned} \tilde{D}^{(k)} &= (b_{w_k} - y_k^* y_k \tilde{b}_{w_k})\tilde{D}^{(k-1)} - (b_{w_k} - \tilde{b}_{w_k}) \\ & \left(\tilde{D}^{(k-1)} u_k\, u_k^* + y_k y_k^* \tilde{D}^{(k-1)} - q^{(k-1)}\, y_k u_k^* + \frac{y_k^* \tilde{D}^{(k-1)} u_k\, \tilde{D}^{(k-1)} - \tilde{D}^{(k-1)} u_k y_k^* \tilde{D}^{(k-1)}}{q^{(k-1)}} \right) \end{aligned} \tag{39}$$

$$q^{(k)} = (b_{w_k} - y_k^* y_k \tilde{b}_{w_k}) q^{(k-1)} - (b_{w_k} - \tilde{b}_{w_k})\, y_k^* \tilde{D}^{(k-1)} u_k, \tag{40}$$

from $\tilde{D}^{(0)} = I_p$ and $q^{(0)} = 1$.

Remark : recall that, if $q^{(k)}$ is the polynomial computed at the k^{th} step of the Schur algorithm, then
(i) $q^{(k)}$ has degree k, as follows from the study of the Schur transform.
(ii) $\tilde{q}^{(k)}(z) = z^k \bar{q}^{(k)}(1/z)$ and $\tilde{D}^{(k)}(z) = z^k\, {}^t\bar{D}^{(k)}(1/z)$.

If the norm of some Schur parameter goes to 1 when integrating the opposite of the gradient, one and only one of the following possibilities occurs :
1) the chart is no more available and we must choose another one,
2) the boundary of $\mathcal{I}_n^p(I)$ is reached.

In practice, if $\|y_k\| \simeq 1$ for some k, we check if $|\tilde{q}^{(k)}(0)|$ is smaller than a fixed small positive number ϵ. If it is, we choose another chart, if not, the boundary is reached. Since divisions by the leading coefficient of $q^{(k)}$ are involved in (39) and (10), this test is particularly adapted to the problem.

7.1 The boundary of $\mathcal{I}_n^p(I)$.

We are going to prove that for any inner matrix Q of degree strictly less that n, we can find a unitary matrix $\mathcal{U}$ and a chart $(\mathcal{V}_n, \phi_n)$ of $\mathcal{I}_n^p(I)$, such that $\mathcal{U}Q$ is a boundary point of $\mathcal{V}_n$ and condition (i) of Proposition 4 is satisfied at each step.

Let $Q \in \mathcal{I}_d^p(I)$ for some $d < n$. Let $\mathbf{y} = (y_1, \dots, y_d)$ its Schur parameters in a chart $(\mathcal{V}_d, \phi_d)$ defined by $\mathbf{w} = (w_1, \dots, w_d)$, and $\mathbf{u} = (u_1, \dots, u_d)$.

The Schur transform $S_{(w,u)}$ applied to some unitary matrix $\mathcal{U}$ and some unit vector y such that $y \neq \mathcal{U}^* u$ will give another unitary matrix. Thus, we can construct a unitary matrix $\mathcal{U}$ in the following way :

$$I_p \to \mathcal{U}_1 = \mathcal{S}_{(w'_1,u'_1)}(I_p, y'_1) \to \mathcal{U}_2 = \mathcal{S}_{(w'_2,u'_2)}(\mathcal{U}_1, y'_2) \to \dots \to \mathcal{U} = \mathcal{S}_{(w'_{n-d},u'_{n-d})}(\mathcal{U}_{n-d-1}, y'_{n-d}),$$

where $w'_1, \dots, w'_{n-d}$, are chosen arbitrarily and $u'_1, \dots, u'_{n-d}, y'_1, \dots, y'_{n-d}$, are unit vectors in $\mathbb{C}^p$, satisfying for $k = 1$ to $n-d$,

$$u_k'^{\,*}\, \mathcal{U}_{k-1}\, y'_k \neq 1, \quad \text{with } \mathcal{U}_0 = I_p.$$

Then, to compute $\mathcal{U}Q$ from the matrix $\mathcal{U}$, we shall use the following property of the Schur transform which can be verified on formula (35) : if $\mathcal{U}$ is a unitary matrix, then

$$S_{(w,\,\mathcal{U}u)}(\mathcal{U}A, y) \;=\; \mathcal{U} S_{(w,\,u)}(A, y).$$

Hence, $\mathcal{U}Q$ can be computed by the following algorithm :

$$I_p \to \mathcal{U}_1 = \mathcal{S}_{(w'_1,u'_1)}(I_p, y'_1) \to \mathcal{U}_2 = \mathcal{S}_{(w'_2,u'_2)}(\mathcal{U}_1, y'_2) \to \dots \to \mathcal{U} = \mathcal{S}_{(w'_{n-d},u'_{n-d})}(\mathcal{U}_{n-d-1}, y'_{n-d})$$

$$\to \mathcal{U}Q^{(1)} = \mathcal{S}_{(w_1,\,\mathcal{U}u_1)}(\mathcal{U}, y_1) \to \mathcal{U}Q^{(2)} = \mathcal{S}_{(w_2,\,\mathcal{U}u_2)}(\mathcal{U}Q^{(1)}, y_2) \dots \to \mathcal{U}Q.$$

Let $\mathbf{w}' = (w'_1, \dots, w'_{n-d}, w_1, \dots, w_d)$, $\mathbf{u}' = (u'_1, \dots, u'_{n-d}, u_1, \dots, u_d)$, and let $(\mathcal{V}_n, \phi_n)$ be the chart of $\mathcal{I}_n^p(I)$ defined by $\mathbf{w}'$,$\mathbf{u}'$. Then $\mathcal{U}Q$ will be considered as a limit point in this chart with Schur parameters $\mathbf{y}' = (y'_1, \dots, y'_{n-d}, y_1, \dots, y_d)$.

Lemma 5 *For $k = 1$ to n, the polynomial $q^{(k)}$ computed from $\mathbf{w}'$,$\mathbf{u}'$ and $\mathbf{y}'$ by formulas* (39) *and* (40) *has degree n, or equivalently $\tilde{q}^{(k)}(0) \neq 0$.*

Proof. We have already proved this property for the usual Schur algorithm, but here, the first few Schur parameters have norm one. However, since $Q^{'}$ may be obtained by a limit process from the usual Schur algorithm, this property is still valid as we shall see.

Indeed, for $k = 1$ to $n - d$, $\|y'_k\| = 1$ and the algorithm builds unitary matrices $\mathcal{U}_1 \dots, \mathcal{U}_{n-d} = \mathcal{U}$. The Schur formula for the polynomial denominator yields

$$\tilde{q}^{(k)} = (\tilde{b}_{w'_k} - b_{w'_k})\tilde{q}^{(k-1)}(1 - {u'_k}^* U_{k-1} y'_k).$$

Since ${u'_k}^* \mathcal{U}_{k-1} y'_k \neq 1$, we get recursively that $\tilde{q}^{(k)}(0) \neq 0$. Moreover, $\tilde{q}^{(k)}$ has degree k, has all its roots on the unit circle and divides $D^{(k)}$, since the matrix $D^{(k)}/\tilde{q}^{(k)}$ is unitary. This is true in particular for $\tilde{q}^{(n-d)}$. But it is clear from the Schur formulas (39) and (40) that $\tilde{q}^{(n-d)}$ keeps on dividing the matrix numerator and polynomial denominator at the subsequent steps. When we reach the step $n_1 = n - d + 1$ we are back to the usual Schur algorithm, that is the norms of the Schur parameters are now strictly less than 1 again. We have

$$\tilde{q}^{(n_1)} = \tilde{q}^{(n-d)}(\tilde{b}_{w_1} - y_1^* y_1 b_{w_1})(1 - u_1^* \mathcal{U} y_1),$$

and more generally, for $k \geq n_1$, $\tilde{q}^{(k)}$ is the product of $\tilde{q}^{(n-d)}$ and the polynomial we should have obtained if the Schur algorithm had started at the step $n - d$. But we know that this one does not vanish at zero (cf remark after (40)), and thus $\tilde{q}^{(k)}(0) \neq 0$. $\square$

7.2 Recursive properties of Ψ_n.

We are now able to prove

Proposition 5 *Assume that $G(z) = {}^t\bar{F}(1/z)/z$ is analytic in $D_r = \{z, \quad \|z\| \leq r\}$ for some $r > 1$.Let $Q \in \mathcal{I}_d^p(I)$ and $Q^{'} = \mathcal{U}Q$ be the boundary point of $\mathcal{I}_n^p(I)$ with coordinates $\mathbf{y}' = (y'_1, \dots, y'_n)$ in $(\mathcal{V}_n, \phi_n)$ (cf. section 7.1).*
Then, $\Psi_{|\mathcal{V}_n}$ extends smoothly in the neighbourhood of $\mathbf{y}'$ to some function denoted by $\overline{\Psi}_{|\mathcal{V}_n}$. Moreover, we have

$$\overline{\Psi}_{|\mathcal{V}_n}(\mathbf{y}') \;=\; \Psi_d(Q).$$

Proof. We first extend the expression of $\tilde{D}$ and q as given by the Schur algorithm, by means of the formulas (39) and (40). To this end, we must observe the two following properties.

Lemma 6 *The coefficients of the polynomials $q^{(k)}$ and $D_{i,j}^{(k)}$ are rational functions of the Schur parameters y_1, y_2,$\ldots y_n$, their denominators being $\tilde{q}^{(1)}(0)\tilde{q}^{(2)}(0)\ldots\tilde{q}^{(k-1)}(0)$.*

Proof. Obvious. □

Lemma 7 *There exists a neighbourhood $\mathcal{W}$ of $\mathbf{y}'$, such that in $\mathcal{W}$ the coefficients of the polynomials $\tilde{D}_{ij}$ and q are bounded rational functions of the components of the Schur parameters. In particular, they depend smoothly on these components and their differentials are bounded in $\mathcal{W}$.*

Proof. This is a consequence of Lemma 5. □

In the sequel, we shall assume that in $\mathcal{W}$, $|\tilde{q}(0)| \geq \mu$, for some $\mu > 0$.

Now, we can extend $\Psi_{|\mathcal{V}_n}$ to $\mathcal{W}$. We shall use the expression of Ψ_n given by Proposition 8. Of course, the polynomial matrices D, R and P, and the polynomial q depend on $\mathbf{y}$. A well-known integral representation for our remainder (cf. [13]) is

$$R(z) \;=\; \frac{1}{2i\pi}\int_{\mathbf{T}} \frac{G\tilde{D}}{q}\frac{q(\xi)-q(z)}{\xi - z}d\xi.$$

If $\mathbf{y}$ belongs to $\mathcal{W}$, the roots of q are no longer in $\mathbf{U}$, but since q depends continuously on $\mathbf{y}$, we can restrict $\mathcal{W}$ so that the roots of q lie in a disk $D_s = \{z, |z| < s\}$ for some s, $1 < s < r$. Let Γ be a contour lying in the open annulus between D_s and D_r. We can extend R in $\mathcal{W}$ by putting

$$R(z) \;=\; \frac{1}{2i\pi}\int_{\Gamma} \frac{G\tilde{D}}{q}\frac{q(\xi)-q(z)}{\xi - z}d\xi. \tag{41}$$

The coefficient of order k of R is given by

$$R_k \;=\; \frac{1}{2i\pi}\int_{\Gamma} \frac{G\tilde{D}}{q}(\xi^{n-k-1}q_n + \ldots + q_{k+1})d\xi.$$

Lemma 7 and the inequality

$$|q(\xi)| > \mu\; d(\Gamma, D_s)^n,$$

imply that the integrand is bounded, and its derivatives (it is smooth) are also bounded. Finally, Lebesgue's theorem say that the integral representation (41) defines a smooth function. The extension of R is still the remainder of the division

$$G\tilde{D} \;=\; V\,q \;+\; R.$$

In $\mathcal{W}$, q keeps on dividing RD and the quotient smoothly extends P. As for R, $\Psi_{|\mathcal{V}_n}$ extends smoothly by the integral representation

$$\Psi_{|\mathcal{V}_n}(Y) \;=\; \|F\|_2^2 - \frac{1}{2i\pi}\mathrm{Tr}\overline{\int_\Gamma G\frac{\tilde{P}}{q}dz}.$$

At $\mathbf{y}'$, the Schur formulas give $(Q^{'})^{-1} = \frac{\tilde{D}^{'}}{q'}$, where q' has degree n. On the other hand, $(Q^{'})^{-1} = Q^{-1}\mathcal{U}^{-1}$, and $Q^{-1} = \tilde{D}/q$, where q has degree d, the two fractions being linked by the relations $q' = uq$ and $\tilde{D}^{'} = u\tilde{D}\mathcal{U}^{-1}$, for some polynomial u having all its roots on the unit circle. It follows from the properties of the division that if $G\tilde{D}^{'} \;=\; V^{'}\; q' \;+\; R^{'}$ and $G\tilde{D} \;=\; V\; q \;+\; R$, then $R^{'} = uR\mathcal{U}^{-1}$; and if $P = RD/q$ and $P^{'} = R^{'}D^{'}/q'$, then $P^{'} = uP$. Finally,

$$\overline{\Psi}_{|\mathcal{V}_n}(\mathbf{y}') = \|F\|_2^2 - < F, \frac{\tilde{P}^{'}}{q'} >= \|F\|_2^2 - < F, \frac{\tilde{P}}{q} >= \Psi_d(Q).$$

□

Let us give two important consequences of Proposition 5.

Lemma 8 *Let $Q \in \mathcal{I}_k^p(I)$ for some $k < n$, $\mathbf{y} = (y_1, \ldots, y_k)$ its Schur parameters in some chart $(\mathcal{V}_k, \phi_k)$ defined by $\mathbf{w} = (w_1, \ldots, w_k)$, and $\mathbf{u} = (u_1, \ldots, u_k)$. Let $w_0 \in \mathbf{U}$, u_0 and y_0 two distinct unit vectors, and*

$$\mathcal{U} = I_p - \frac{(u_0 - y_0)(u_0 - y_0)^*}{1 - u_0^* y_0}.$$

Then, $Q^{'} = \mathcal{U}Q$ is a boundary point of $\mathcal{V}_{k+1}$ defined by $\mathbf{w}' = (w_0, w_1, \ldots, w_k)$, and $\mathbf{u}' = (u_0, \mathcal{U}u_1, \ldots, \mathcal{U}u_k)$, and $\mathbf{y}' = (y_0, y_1, \ldots, y_k)$ are its Schur parameters. Moreover,

$$\overline{\Psi}_{|\mathcal{V}_{k+1}}(\mathbf{y}') = \Psi_{|\mathcal{V}_k}(\mathbf{y}).$$

Corollary 1 *Suppose that $\mathbf{y}$ is a local minimum of $\Psi_{|\mathcal{V}_k}$. Then, the gradient of $\overline{\Psi}_{|\mathcal{V}_{k+1}}$ at $\mathbf{y}'$ is orthogonal to the surface $\mathcal{S} = \{(y_0, \ldots, y_n),\ \|y_0\| = 1,\ \|y_j\| < 1,\ j = 1, \ldots n\}$ and points outwards.*

Proof. From Proposition 5 we see that the projection of the gradient of $\overline{\Psi}_{|\mathcal{V}_{k+1}}$ at $\mathbf{y}'$ on $\mathcal{S}$ is just the gradient of $\Psi_{|\mathcal{V}_k}$ at $\mathbf{y}$, whence orthogonality holds. Moreover, it cannot point inwards because this would imply that $Q^{'}$ which is rational of order k is a local minimum at order $k+1$, and this is impossible except if F itself has degree $k+1$ (cf. [2]). □

We are now able to describe our algorithm.

7.3 The algorithm.

Recall that $G(z) = {}^t\bar{F}(1/z)/z$ is assumed to be analytic in $D_r = \{z, \quad \|z\| \leq r\}$ for some $r > 1$. We shall make two extra assumptions in what follows. First, we shall assume that grad Ψ_k does not vanish on the boundary of $\mathcal{I}_k^p(I)$, for $1 \leq k \leq n$. Second, we shall require all the critical points of Ψ_k in $\mathcal{I}_k^p(I)$ to be nondegenerate, i.e. to have a second derivative which is a nondegenerate quadratic form. These two properties hold generically, that is for almost every F in some sense, and we refer the reader to [3] for the first one, and [2] for the second one. They ensure in particular that critical points in $\mathcal{I}_k^p(I)$ are finite in number.

Taking this for granted, we are now able to describe a procedure to determine a *local* minimum of Ψ_n. The algorithm proceeds as follow.

Choose some $\mathbf{y}_0$ in some chart $\mathcal{V}_n$ of $\mathcal{I}_n^p(I)$ as an initial condition, and integrate the vector field -grad Ψ_n. There are two possibilities : either we reach a critical point or the norm of some Schur parameter goes to 1. In the first case, if the critical point is a minimum, we are done. Otherwise, since it is nondegenerate, the critical point will be unstable under small perturbations, thereby allowing us to continue the procedure. Since Ψ_n decreases, we cannot meet the same critical point twice, and because such points are finite in number, we succeed unless the norm of some Schur parameter goes to 1.

If the norm of some Schur parameter goes to 1, there are still two possibilities : either the chart is no more available (condition (ii) of Proposition 4 is satisfied at some step of the Schur algorithm) and we must choose another one, or we have reached some boundary point Q_d of degree d, with $d < n$. In this case, we first multiply Q_d by a unitary matrix in order to satisfy (3), then we look for some chart in which Q_d can be represented, and finally we compute its Schur parameters. The degree d is nonzero, since Ψ_0 is a constant function whose value $\|F\|_2^2$ is the maximum of Ψ_n on $\mathcal{I}_n^p(I)$. So, we can begin all over again replacing n by d in the case of a boundary point. We are bound to reach a local minimum of Ψ_k for some k satisfying $1 \leq k < n$, at some point of $\mathcal{I}_k^p(I)$.

Starting from this point, we construct a boundary point of $\mathcal{I}_{k+1}^p(I)$ as in Lemma 8. By Lemma 1, the gradient of $\overline{\Psi}_{|\mathcal{V}_{k+1}}$ at this boundary point is orthogonal to the surface $\mathcal{S} = \{(y_0, \ldots, y_n), \; \|y_0\| = 1, \; \|y_j\| < 1, \; j = 1, \ldots n\}$ and points outwards. Therefore, integrating -grad $\overline{\Psi}_{|\mathcal{V}_{k+1}}$ from this point leads us to penetrate into the interior of $\mathcal{I}_{k+1}^p(I)$, so that the whole process can be carried over again, with n replaced by $k+1$.

Since the value of the criterion decreases continuously, we never meet twice the same critical point and this ensures that the procedure eventually comes to an end.

8 Numerical examples.

We have run a number of examples using the software package Scilab. In particular, it contains the function *ode* for computing the solution of a system of differential equations. The purpose of the following is to demonstrate the procedure of computing a local minimum.

Let

$$F(z) = \begin{pmatrix} \frac{1+z}{z^2-z+1/4} & \frac{1}{z-1/2} \\ \frac{-z^2+z+1}{z^3+1/2z^2-1/4z-1/8} & \frac{z-1/4}{z^2+z+1/4} \end{pmatrix},$$

or else $F = N/d$, where

$$N(z) = \begin{pmatrix} z^3+2z^2+5/4z+1/4 & z^3+1/2z^2-1/4z-1/8 \\ -z^3+3/2z^2+1/2z-1/2 & z^3-5/4z^2+1/2z-1/16 \end{pmatrix},$$

and

$$d(z) = z^4 - 1/2z^2 + 1/16.$$

The function F is given, as input of the program, by the 200 first Fourier coefficients of its rational entries.

We present several ways of approximating this 4th order system at its own order from different starting points. Of course, we must recover the function F itself, and this situation is particularly simple since from consistency, the criterion has no other critical points.

We have noticed (cf. Section 5) that one choice for w is enough to describe $I_n^p(I)$. In the following, we shall work with a single value for w : $w = 0$. Thus, a point is given by a chart's parameter $\mathbf{u}$ and a Schur parameter $\mathbf{y}$.

1) We first start from the initial point parametrized by $\mathbf{y}$ in the chart given by $\mathbf{w}, \mathbf{u}$:

$\mathbf{u}$	1 0	1 0	0 1	1 0
$\mathbf{y}$	0.5 −0.5	−0.5 −0.5	0.5 0.5	0.5 −0.5

At this point, the criterion is equal to 7.3324036. We get a minimum in this chart for

$\mathbf{y}_{min}$	0.4947911 −0.3200608	−0.5709638 0.3282461	0.2658280 0.2025660	0.1465519 −0.1293104

with a criterion less than 10^{-8}. The best approximant computed from these parameters agrees with F up to 6 decimal places.

2) An example in which the boundary is reached.

Step 1. The starting point is given by

u	1	1	0	1
	0	0	1	0
y	0.5	0.5	−0.5	0.5
	0.5	0.5	−0.5	0.5

and the criterion is equal to 9.1099142. We meet the boundary at

$\mathbf{y}_{bound}$	0.5093948	0.3573557	−0.6598669	0.5560878
	0.8605213	0.5500306	−0.4055843	0.2648168

,

where the first parameter has a norm nearly equal to 1, while $\tilde{q}^{(1)}(0) = 0.4906052$ stays away from 0. The criterion is equal to 3.7862344.

Step 2. This point is converted into an initial point for a 3rd order approximation, with the same criterion, applying Lemma 8 with $u_0 = {}^t(1\ 0)$ and $y_0 = {}^t(0.5093948\ 0.8605213)$:

u	0.5093948	0.8605213	0.5093948
	0.8605213	−0.5093948	0.8605213
y	0.3573557	−0.6598669	0.5560878
	0.5500306	−0.4055843	0.2648168

.

We find a 3rd degree minimum for

$\mathbf{y}_{min}$	−0.5744981	0.0214156	0.2054699
	0.6522901	−0.4332472	0.4280223

,

where the criterion is equal to 0.9973472 and the relative error to 0.0507508.

Step 3. This point provides starting points for 4th order integrations. For instance, applying Lemma 8 with $u_0 = {}^t(1\ 0)$ and $y_0 = {}^t(0\ 1)$, let us start from

u	1	0.8605213	−0.5093948	0.8605213
	0	0.5093948	0.8605213	0.5093948
y	0	−0.5744981	0.0214156	0.2054699
	1	0.6522901	−0.4332472	0.4280223

.

The integration leads to some point for which the first parameter has a norm equal to 0.9870436 and $\tilde{q}^{(1)}(0) = 0.0136804$. The approximant computed from this point agrees with F with only 2 significant digits. The low value of $\tilde{q}^{(1)}(0) = 0.0136804$ can explain the poor agreement of the results, and in this case, we had better choose another chart (cf. Proposition 4 and the following remark). We have observed that the results become inacurate when some $\tilde{q}^{(k)}$ becomes less than 0.125.

Step 4. Restart the integration. At

$\mathbf{y}_{min}$	0.8796394 0.0962381	−0.6880990 0.1018120	0.1692177 0.232071	0.2642769 −0.0267232

,

the value of $\tilde{q}(0)$ is less than 0.125, and this could induce an important error in the computation. It is better to find another chart. To this end, we choose at each step of the Schur algorithm, u_k among ${}^t(1\ 0)$ and ${}^t(0\ 1)$ in order to get the smallest Schur parameter. Our point in this new chart is given by

$\mathbf{u}$	1 0	1 0	0 1	1 0
$\mathbf{y}$	.6316952259 −.2779806122	−.5785967337 −.3369857917	.2621593118 .1922908023	.1566999089 −.1421830569

.

Step 5. The integration can continue, and the minimum is reached for

$\mathbf{y}_{min}$	0.4951705 −0.3199460	−0.5709872 −0.3282689	0.2658132 0.2025562	0.1465757 −0.1293524

.

The approximant computed from these parameters agrees with F with 4 significant digits. This point can be compared with the result of example 1 in the same chart.

3) In many real datas examples, the choice of the initial point is very important. In a great number of cases, the computation leads to a local minimum and not the global one. The following procedure has a good chance to provide the global minimum.

Order 1. We start from the boundary point

$\mathbf{w}$	0
$\mathbf{u}$	1 0
$\mathbf{y}$	−1 0

and we get a first order minimum at

$\mathbf{y}_1$	−0.7739451 −0.1598369

,

where the criterion is equal to 9.553057 and the relative error to 0.4861148.

Order 2. This minimum provides a starting point with the same criterion for a second order approximation :

w	0	0
u	1	0
	0	1
y	0	−0.7739451
	1	−0.1598369

.

We get a minimum at

$\mathbf{y_2}$	−0.7182560	−0.4167791
	−0.2769471	0.6720201

,

where the criterion is equal to 2.5090541 and the relative error to 0.1276752.

Order 3. This minimum provides a starting point for a 3rd order approximation :

w	0	0	0
u	1	−1	0
	0	0	1
y	−1	−0.7182560	−0.4167791
	0	−0.2769471	0.6720201

.

We get a minimum at

y	−0.6264395	−0.2041812	0.0900264
	−0.2980447	0.1589819	0.6333477

,

where the criterion is equal to 0.9974956 and the relative error to 0.0507584.

Order 4. This minimum provides a starting point for a 4th order approximation :

w	0	0	0	0
u	1	0	0	−1
	0	−1	1	0
y	0	−0.6264395	−0.2041812	0.0900264
	−1	−0.2980447	0.1589819	0.6333477

which leads to the minimum with a very good accuracy.

9 Conclusion

We have described here a gradient algorithm which nearly solved our initial matrix rational approximation problem, problem useful in the framework of identification of linear systems :

"given F, a $p \times m$ stable transfer function, find a rational stable function H of degree at most n which minimizes the L^2-norm of $F - H$". We have proved that our algorithm converges to a local minimum of the L^2 criterion. This is to our knowledge, the first algorithm of this type.

We have implemented this algorithm using the matrix-based scientific software Scilab. This choice was very well adapted to improve our method. The algorithm succeeded in many simple examples demonstrating the procedure of computing a local minimum. Concerning the numerical aspects of the algorithm, it seems that our implementation does not offer a sufficient computation speed, and computer programs are being written to test the algorithms on real-data. For instance, with our prototype, considering an identification problem of hyperfrequency filters we have only found local minima of order 8.

If we want to solve the whole initial problem, we must now answer this question : "how can we reach the global minimum ?". It is possible to consider the difficulty as intrinsic and cope with it, trying to find the global minimum at any cost. One may think of initializing the algorithm at enough points to reach all local minima and compare between them. But we do not know what "enough" means and we do not have a bound for the number of initializing points. Consequently, more efficient strategies should be investigated. For instance, we can try to find as many local minima of order $k-1$ as we can to initialize our procedure at order k, by replacing them on the boundary of $\mathcal{I}_k^p(I)$, as we have shown in Lemma 8. We begin with $k = 1$, and so on, step by step, until $k = n$. But even though, we are not sure to find the global minimum, and this method is very slow . Except for very special functions, for instance the case of functions which have only one minimum, we do not know a better method. However, in many problems, we hope that some more information or engineering judgement, could help us to select an initial point which ensures rapid convergence of the procedure to the global minimum.

An interesting and useful extension of our work could be to allow F to have poles both inside and outside the unit disk. We can think about considering unitary matrices instead of inner matrices. That could be done by allowing the norms of the Schur parameters to be greater than one, using an extension of our "polynomial tangential Schur algorithm".

Aknowledgements : the authors would like to thank D. Alpay who introduced them to Schur analysis and L. Baratchart for providing them with several helpful suggestions during this research. We also like to thank E. Saff and A. Gombani for their very careful reading of the first version of this paper and useful criticisms.

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, $n^0 2$ (1991), pp. 413–418.

[4] L. Baratchart and M. Olivi, *Inner-unstable factorization of stable rational transfer functions*, in Modeling, Estimation and Control of Systems with Uncertainty, A. K. G.B. Di Masi, A. Gombani, ed., vol. 10 of Progress in System and Control Theory, Birkhauser, 1991, pp. 22–39.

[5] L. de Branges and J. Rovniyak, *Canonical models in quantum scattering theory*, in C. Wilcox, editor, Perturbation theory and its applications in quantum mechanics, Holt, Rinehart and Winstom - New York, 1966.

[6] 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.

[7] P. Fuhrmann, *Linear systems and operators in Hilbert spaces*, Mc Graw Hill, 1981.

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

[9] T. Kailath, *Linear Systems*, Prentice-Hall, 1980.

[10] L. Meier, *Approximation of linear constant system*, in I.E.E. Trans. Automat. Contr. AC-12, 1967, pp. 585–588.

[11] V. Potapov, *The multiplicative structure of j-contractive matrix functions*, American mathematical society translations, 15 (1960).

[12] L. Schwartz, *Sous–espaces hilbertiens d'espaces vectoriels topologiques et noyaux associés (noyaux reproduisants)*, Journal d'analyse mathématique, (1964), pp. 115–256.

[13] J. L. Walsh, *Interpolation and Approximation by rational Functions in the complex Domain*, A.M.S. Publications, 1962.

Unité de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unité de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unité de recherche INRIA Rhône-Alpes, 46 avenue Félix Viallet, 38031 GRENOBLE Cedex 1
Unité de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unité de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399