Identification and rational L^2 approximation: a gradient algorithm.

Laurent Baratchart, Michel Cardelli, Martine Olivi [1]

Abstract: This paper deals with the identification of linear constant dynamical systems when formalized as a rational approximation problem. The criterion is the l^2 norm of the transfer function, which is of interest in a stochastic context. The problem can be expressed as nonlinear optimization in a Hilbert space, but standard algorithms are usually not well adapted. Here, we present a generic recursive procedure to find a local optimum of the criterion in the case of scalar systems. Our methods are borrowed from differential theory mixed with a bit of classical complex analysis. To our knowledge, the algorithm described in this paper is the first that ensures convergence to a local minimum. Finally, we discuss a number of unsettled issues.

1 Introduction.

In this paper, we approach the problem of identification within the framework of Hardy spaces by considering this as a rational approximation problem. We restrict ourselves to linear constant strictly causal single-input single-output dynamical systems (in short systems). We first consider the case of a discrete time system. Let $f_1, f_2, ..., f_m, ...$ be its impulse response. Identifying the system usually means recognizing the sequence (f_m) as the Taylor coefficients at infinity of a proper rational function whose denominator degree (in irreducible form) is then the order of the system. But since such a sequence might not exist, in practice one has to be content with finding a rational sequence (r_m) that resembles (f_m). This, of course, has no definite meaning, and some criteria has to be chosen. Such criteria can occur in connection with stability. Assume, for instance, that the system is l^k-stable for some $k \geq 1$, that is

$$\sum_{m=1}^{\infty} |f_m|^k < \infty.$$

[1]Institut National de Recherche en Informatique et Automatique, route des Lucioles, Sophia-Antipolis 06565 Valbonne Cedex (France)

One can then look for some (r_m) which is close to (f_m) in the l^k sense. Since increasing the order of (r_m) arbitrarily is unacceptable, it is also reasonable to bound it from above by some number n. Note that the identified model (r_m) will then automatically be stable.

Our assumptions have as an effect that the transfer function

$$f(z) = \sum_{m=1}^{\infty} f_m z^{-m}, \quad \text{as well as the rational function} \quad r(z) = \sum_{m=1}^{\infty} r_m z^{-m}$$

are holomorphic for $|z| > 1$, and one can ask in which sense they are close to each other from an analytic viewpoint. In general, there is no completely satisfactory answer. A partial result in this direction involves the so-called real Hardy spaces H_μ^-, where $1 \le \mu \le \infty$, that we now define. If h is analytic for $|z| > 1$ and $\rho > 1$ is a real number, define a function h_ρ on the unit circle T by putting $h_\rho(e^{i\theta}) = h(\rho e^{i\theta})$. By definition, H_μ^- will consist of those h vanishing at infinity, assuming real values for real arguments and such that

$$\sup_{\rho>1} \|h_\rho\|_\mu < \infty,$$

where $\| \ \|_\mu$ is the norm in $L^\mu(T)$. It is then standard to show (FUHRMANN, 81, th. 12.11) that h has a radial limit h^* almost everywhere on T which lies in $L^\mu(T)$, whose Fourier coefficients are real and those of non-negative rank do vanish. Moreover, we have $\|h^*\|_\mu = \sup_{\rho>1} \|h_\rho\|_\mu$. Conversely, any member of $L^\mu(T)$ with Fourier coefficients as above is the radial limit of some unique element of H_μ^-. One can then identify h and h^* and consider $\| \ \|_\mu$ as a norm on H_μ^-, that defines it as a Banach space.

Now, if k' is the conjugate of k ($1/k + 1/k' = 1$), the Hausdorff-Young theorem (DUREN 70, th. 6.1) gives

- when $1 \le k \le 2$, $$\|f\|_{k'} \le \left(\sum_{m=1}^{\infty} |f_m|^k \right)^{1/k}, \tag{1}$$

so that f belongs to $H_{k'}^-$ whenever (f_m) belongs to l^k, and that to be close in the l^k sense for impulse responses implies to be close in the $H_{k'}^-$ sense for transfer functions;

- when $2 \le k \le \infty$, $$\left(\sum_{m=1}^{\infty} |f_m|^k \right)^{1/k} \le \|f\|_{k'}, \tag{2}$$

so that to be close in the $H_{k'}^-$ sense for transfer functions implies to be close in the l_k sense for impulse responses.

Of course, there may be other reasons for being interested in rational approximation, even if $2 < k'$. The H_∞^- norm, for instance, is the operator norm $l^2 \to l^2$. Nevertheless, rational approximation in $H_{k'}^-$ is demonstrably relevant to identification of impulse responses only if $1 \leq k' \leq 2$.

Clearly from the above, the case where $k = k' = 2$ is particularly nice. The conjunction of (1) and (2) is just Parseval's equality:

$$\|f\|_2 = \left(\sum_{k=1}^{\infty} |f_k|^2 \right)^{1/2},$$

and H_2^- is a Hilbert space with scalar product

$$< f, h >= \frac{1}{2\pi} \int_0^{2\pi} f(e^{i\theta}) \, \overline{h(e^{i\theta})} \, d\theta$$

which in turn can be converted into a line integral

$$< f, h >= \frac{1}{2i\pi} \int_T f(z) \, h(\frac{1}{z}) \, \frac{dz}{z}.$$

In the rest of this paper, we shall restrict ourselves to rational approximation in H_2^- which we phrase as follows:

For $f \in H_2^-$, minimize $\|f - r\|_2$ where r ranges over all rational functions in H_2^- of order at most n.

Note that a rational fraction $r = p/q$ belongs to H_2^- if and only if $deg(p) < deg(q)$ and the roots of q lie inside the unit disk U.

There is a probabilistic interpretation of this criterion in identification: if f is the transfer function of a l^2 stable system driven by a white noise δ, the output $y = f\delta$ is a stationary process. If the latter is to be modelled by a rational function p/q of order at most n, and if we put $\hat{y} = p/q\,\delta$, the minimization of the covariance of $y - \hat{y}$ is achieved when $\|f - p/q\|_2$ is itself minimal.

In this paper, we present an algorithm to find *local* best approximants of a given order n, which proceeds recursively by numerically solving differential equations over a *compact* subset of $\mathbf{R}^n$. This procedure is the first one, to our knowledge, for which convergence is guaranteed, at least generically. We then present a convincing experiment, and we finally list some open questions. Most proofs are just sketched, since our main concern here is not technical but rather to describe the procedure.

Let us first list a few known results concerning this question. It can be proved (see e.g., (BARATCHART, 1986), (RUCKEBUSCH, 1978) or (WALSH, 1962)) that the problem stated above has a solution. This solution is not always unique, but generically it is (BARATCHART, 1987), though there might be lots of local minima. One can show that if f is not a rational function of order less than n, a case which will be implicitly ruled out in what follows, no local minimum can be of order less than n (RUCKEBUSCH, 1978). In other words, one should always take advantage of all parameters at hand.

This observation leads to the conclusion that *it is enough to minimize the norm over the set of irreducible fractions of H_2^- of order exactly n.* Since this set is a manifold, it is possible to use classical tools from optimization, like steepest descent algorithms. However, due to the shape of the gradient vector field and to the non-compactness of the domain over which we optimize, these methods may fail to converge. In the remaining, we develop a different approach, which is based on the elimination of some parameters, and gives rise to a much nicer geometric picture.

2 The function Ψ_n.

Let P_n be the set of real polynomials of degree at most n, and $\mathcal{P}_n^r$ the subset of monic polynomials of degree n whose roots are in the disk U_r of radius r.

$$\text{We look for} \qquad \min_{p,q} \|\frac{p}{q} - f\|_2^2, \tag{3}$$

where $p \in P_{n-1}$ and $q \in \mathcal{P}_n^1$. Consider the n-dimensional linear subspace of H_2^- defined by $V_q = P_{n-1}/q$. For fixed q, the minimum in (3) is obtained when p/q is the orthogonal projection $\pi_q(f)$ of f onto V_q. If we define a polynomial $L_n(q) \in P_{n-1}$ by the formula $L_n(q) = q\pi_q(f)$, we are thus led to minimize the function

$$\Psi_n \ : \ \mathcal{P}_n^1 \to \mathbf{R} \ \text{ defined by } \ \Psi_n(q) = \left\| f - \frac{L_n(q)}{q} \right\|_2^2.$$

The polynomial $L_n(q)$ must satisfy by definition

$$\forall j = 0, .., n-1, \qquad < f - \frac{L_n(q)}{q}, \frac{z^j}{q} >= 0,$$

that is to say

$$\frac{1}{2i\pi} \int_T (f(\frac{1}{z}) - \frac{L_n(q)}{q}(\frac{1}{z}))\ \frac{u}{q}(z)\ \frac{dz}{z} = 0,$$

for any complex polynomial u of degree at most $n-1$.

Let us define the function g holomorphic in U by putting

$$g(z) = f(1/z)/z. \tag{4}$$

Define further $\tilde{q} \in P_n$ as $z^n q(1/z)$. The roots, possibly at infinity, of $\tilde{q}$ are the inverses of those of q. Similarly, we shall also denote by $\widetilde{L_n(q)}$ the polynomial $z^{n-1} L_n(q)(1/z)$.

With these notations, our integral equation becomes

$$\frac{1}{2i\pi} \int_T (g - \frac{\widetilde{L_n(q)}}{\tilde{q}}) \frac{u}{q}\, dz = 0,$$

whenever u is a complex polynomial of degree at most $n-1$. Since $g - \widetilde{L_n(q)}/\tilde{q}$ belongs to the Hardy space $H_2(U)$ of the unit disk U (DUREN,1970), the residue theorem applies giving that the above is satisfied if and only if $g - \widetilde{L_n(q)}/\tilde{q}$ matches zero at each root of q, counting multiplicities. Namely q should divide $g\tilde{q} - \widetilde{L_n(q)}$. Thus, $\widetilde{L_n(q)}$ is the unique polynomial of degree at most $n-1$ interpolating $g\tilde{q}$ at the zeroes of q, that is to say:

Proposition 2.1 *The polynomial $\widetilde{L_n(q)}$ is the remainder of the division of $g\tilde{q}$ by q.*

A well-known integral representation for our remainder (WALSH, 1962, §3.1) is

$$\widetilde{L_n(q)} = \frac{1}{2i\pi} \int_\Gamma \frac{\tilde{q}g(\xi)}{q(\xi)} \left[\frac{q(\xi) - q(z)}{\xi - z} \right] d\xi, \tag{5}$$

where Γ is any contour contained in the domain of holomorphy of g that encompasses the zeroes of q. As usual, the independence of the integral from the contour follows from Cauchy theorem.

3 Extension of the domain of Ψ_n. Smoothness.

A monic polynomial of degree n, $q(z) = z^n + a_{n-1}z^{n-1} + \cdots + a_0$ can be identified with the vector $(a_{n-1}, a_{n-2}, .., a_0)$ of $\mathbf{R}^n$, and this allows for $\mathcal{P}_n^1$ to be considered as an open subset of $\mathbf{R}^n$.

So far, Ψ_n has been defined only on $\mathcal{P}_n^1$. Now, if we assume that f is holomorphic not only for $|z| > 1$ but also in a neighborhood of the unit circle T, we shall be able to extend Ψ_n to a smooth function defined on a neighborhood of the closure, in $\mathbf{R}^n$, of $\mathcal{P}_n^1$. This closure will be denoted by Δ_n, and clearly consists of all real monic polynomials of degree n whose roots are in the closed unit disk $\bar{U}$. Our hypothesis on f, *which will be assumed hereafter*, is equivalent to requiring that g be analytic in the disk U_r of radius r for some $r > 1$ that will remain fixed in the sequel.

Proposition 3.1 *The map Ψ_n extends to a smooth function $\Psi_n : \mathcal{P}_n^r \to \mathbf{R}$.*

Proof: If $q \in \mathcal{P}_n^1$, the properties of the orthogonal projection show that

$$\Psi_n(q) = \left\| f - \frac{L_n(q)}{q} \right\|_2^2 =< f, f > - < f, \frac{L_n(q)}{q} > . \tag{6}$$

Since the contour Γ may be deformed within the domain of holomorphy of g so as to surround any n-tuple of points in U_r, the integral representation (5) obviously yields a smooth extension of L_n to a map $\mathcal{P}_n^r \to P_{n-1}$. Note that $\widetilde{L_n(q)}$ is still the remainder of the division

$$g\tilde{q} = vq + \widetilde{L_n(q)}. \tag{7}$$

Having this at our disposal, it is now sufficient by (6) to extend smoothly for every k, $q \to< f, z^k/q >$ to $\mathcal{P}_n^r$. This can be done similarly by putting

$$< f, \frac{z^k}{q} >= \frac{1}{2i\pi} \int_\Gamma g(\xi) \frac{z^k}{q(\xi)} d\xi,$$

which again allows for q to lie in $\mathcal{P}_n^r$. Q.E.D.

We now turn to Δ_n. It is plain to see that Δ_0 consists solely of the point 0, and Δ_1 of the segment $[-1, 1]$. It is easy to check that Δ_2 is the triangle with vertices $(-2, 1)$,$(2, 1)$ and $(0, -1)$ (see fig.1). It is more difficult to see what Δ_3 looks like (see fig. 2). In general, Δ_n is topologically a ball, as is proved in (BARATCHART, OLIVI, 1988), and we shall concentrate here on its boundary $\partial\Delta_n$, which is homeomorphic to a sphere. This set consists of polynomials in Δ_n having at least one root of modulus 1. Among such polynomials, one can distinguish between those having ± 1 as a root and those having no real root of modulus 1, but which are divisible by some polynomial of the form $z^2 + 2xz + 1$ with $x \in (-1, 1)$. This means that $\partial\Delta_n$ is the union of $(z-1)\Delta_{n-1}$, $(z+1)\Delta_{n-1}$, and $(z^2+2xz+1)\Delta_{n-2}$ for $x \in (-1, 1)$. In other words, $\partial\Delta_n$ is made from two copies of Δ_{n-1} and infinitely many copies of

Δ_{n-2}. As is already apparent when n equals 2 or 3, the boundary $\partial\Delta_n$ is not smooth, namely there are corners. The smooth part of $\partial\Delta_n$ may be characterized as the set of polynomials in Δ_n having exactly *one* irreducible factor over $\mathbf{R}$ with roots of modulus one. Alternatively, they are interior points of the copies of Δ_{n-1} and Δ_{n-2} introduced above. The crux of the matter is the following lemma describing the behaviour of Ψ_n on $\partial\Delta_n$.

Lemme 3.1 *Let $q \in \partial\Delta_n$, and suppose $q = q_u q_i$ where q_u is monic of degree k and has all its roots of modulus 1 while q_i is interior to Δ_{n-k}. Then $\Psi_n(q) = \Psi_{n-k}(q_i)$.*

Proof: From (6), it is sufficient to prove that $L_n(q) = q_u L_{n-k}(q_i)$. But, since inverse and conjugate agree on T, we have $\tilde{q}_u = \pm q_u$ and the result follows from (7). Q.E.D.

Let us denote by $\nabla_n(q)$ the gradient vector of Ψ_n at the point q. Later we will use the following consequence of lemma 1.

Corollaire 3.1 *Let q, as in the previous lemma, belong to the smooth part of $\partial\Delta_n$, and be such that q_i is a critical point of Δ_{n-k} (note that $k = 1$ or 2). Then $\nabla_n(q)$ is orthogonal to $\partial\Delta_n$ and points outwards.*

Proof: From Lemma 1, we see that the projection of $\nabla_n(q)$ on $\partial\Delta_n$ is just $\nabla_{n-k}(q_i)$, so that $\nabla_n(q)$ is orthogonal to $\partial\Delta_n$. Moreover, it cannot point inwards because this would imply that $L_{n-k}(q_{n-k})/q_{n-k}$, which is rational of order $n-k$, is locally a best approximant to f among rational functions of order n, hence that f itself is rational of order $< n$. Q.E.D.

4 A generic algorithm to find a local minimum.

As we said before, the minimum value of Ψ_n on Δ_n can only be taken at some interior point of Δ_n since f is not rational of order less than n by hypothesis. As a consequence, such a point q must be a critical point of Ψ_n, namely has to satisfy

$$\nabla_n(q) = 0.$$

We shall make two extra assumptions in what follows. First, we shall assume that ∇_k does not vanish on $\partial\Delta_k$, for $1 \leq k \leq n$. Second, k given as above, we shall ask all critical points of Ψ_k in Δ_k to be nondegenerate, *i.e.*, to have a second derivative that 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 (BARATCHART, OLIVI, 1988) for a precise statement. They ensure in particular that critical points in Δ_k are finite in number.

Taking this for granted, we are now able to describe a procedure to determine a *local* minimum of Ψ_n. We first define one more bit of notation: if $q \in \Delta_k$ for some $k \leq n$, we define $\Psi(q)$ to be simply $\Psi_k(q)$. This new notation enables us to compare $q \in \Delta_k$ and $q' \in \Delta_{k'}$, but we shall still use Ψ_k when referring to the behaviour on Δ_k only.

The algorithm proceeds as follows.

Choose some interior point q_0 of Δ_n as an initial condition, and integrate the vector field $-\nabla_n$. There are two possibilities: either we reach a critical point or we reach $\partial\Delta_n$. In the former case, if the critical point is a local 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 eventually success or we reach $\partial\Delta_n$.

If we meet $q_b \in \partial\Delta_n$ (see fig. 3), we decompose it as $q_u q_i$ where q_u has all its roots of modulus 1 and q_i has none. From lemma 1, we see that $\Psi(q) = \Psi(q_i)$. Moreover, the degree k of q_i is nonzero, for Ψ_0 is a constant function whose value $\|f\|_2^2$ is the maximum of Ψ_n on Δ_n.

Now, q_i lies in the interior of Δ_k, so we can begin all over again with n replaced by k and q_0 by q_i. Since k decreases but remains strictly positive, we are bound to reach a local minimum of Ψ_m for some m satisfying $1 \leq m < n$, at some interior point q_m of Δ_m.

Consider now the point $q_1 = q_m(z+1)$ of $\partial\Delta_{m+1}$. It lies in a smooth region of the boundary, so that $-\nabla_{m+1}(q_1)$ is orthogonal to $\partial\Delta_{m+1}$, and points inwards by the corollary to lemma 1. Therefore, integrating $-\nabla_{m+1}$ starting from q_1 (see fig. 3) leads us to penetrate into the interior of Δ_{m+1}, so that the whole process can be carried over again, with n replaced by $m+1$. Since Ψ decreases continuously, we never meet twice the same critical point of Ψ_k for $1 \leq k \leq n$, and this ensures that the procedure eventually comes to an end. An end means precisely that we have reached the desired local minimum of

Ψ_n.

5 The continuous time case.

So far, we have only dealt with rational approximation in the Hardy space of the disk, as related to the identification of discrete-time l^2-stable transfer functions. In practice however, continuous-time systems mainly occur. A treatment similar to that in section 1 could be given, but we just indicate briefly how the above technique applies. Let us call a continuous-time system L^2-stable if its impulse response is in $L^2[0,\infty]$. The Paley-Wiener theorem (see e.g., DUREN,70) asserts that the Laplace transform is an isometry between $L^2[0,\infty]$ and the real Hardy space $\mathcal{H}_2$ of right-half plane consisting of functions F holomorphic for $\Re(z) > 0$, satisfying the realness condition $F(\bar{z}) = \bar{F}(z)$, and such that

$$\sup_{x>0} \int_{-\infty}^{+\infty} |F(x+iy)|^2\, dy \ < \ \infty. \tag{8}$$

Here, the squared norm $\|F\|^2_{\mathcal{H}_2}$ is the supremum in (8) by definition. In other words, the set of transfer functions of L^2-stable systems is precisely $\mathcal{H}_2$. The rational approximation problem in this context consists in looking for

$$\min_{p,q} \|F - \frac{p}{q}\|_{\mathcal{H}_2},$$

where p/q ranges over all rational functions of order $\leq n$ in $\mathcal{H}_2$. Note that such fractions are exactly transfer functions of stable systems of order at most n.

A possible interpretation of the L^2 criterion is as follows. Assume an L^2 system is driven by a white noise, so that the output is a stationary stochastic process

$$\xi(t) = \int_{-\infty}^{+\infty} h(t-\tau)\, dW(\tau)$$

where W is a Wiener process. The variance of ξ, which is independent of t, is equal to $\|h\|_2^2$ (DOOB, 1953). If h_n is the impulse response of a system of order at most n, and p/q its transfer function, and if we put

$$\xi_n(t) = \int_{-\infty}^{+\infty} h_n(t-\tau)\, dW(\tau),$$

the variance of $\xi - \xi_n$ is $\|h - h_n\|_2^2$, which is also equal to $\|F - p/q\|^2_{\mathcal{H}_2}$, where F and p/q are respectively the Laplace transforms of h and h_n. Therefore, if we want a model of order at most n for the process, we minimize the covariance of the error by solving the above rational approximation problem.

Now, the question comes back to the one investigated above. Indeed, if we put

$$\varphi : z \to \frac{z+1}{z-1} \quad \text{and} \quad \Phi(F)(z) = 2\sqrt{\pi}\,\frac{F \circ \varphi(z)}{z-1},$$

it turns out that Φ is an isometry between $\mathcal{H}_2$ and H_2^-, and it is plain to see that Φ preserves rational functions and their order. Therefore, putting $f = \Phi(F)$ brings us back to the discrete-time case.

6 Numerical examples.

Until now, we have assumed that the transfer function of the system under consideration exists and is perfectly known to us. But in practice, of course, this is never so. The system certainly exists, but the transfer function may not be defined. And even if it does, it is only known to us through a finite number of experiments, whereas the function f to be approximated in H_2^- has to be completely defined if we actually want to run the algorithm. In fact, the definition of f from a set of experimental data is an arduous problem that was entirely bypassed in the above development.

For instance, if the transfer function F of a continuous-time system is computed through frequency response experiments, we are given its value at a certain number of points of the imaginary axis, so that f is only known at a finite number of points on the unit circle. To estimate its Fourier coefficients, the best we can hope for, in general, is to have a convergent procedure to estimate $\Phi(F)$ when the number of experiments increases. This is in the style of (BARATCHART, 1989), however, we shall not go into further details here.

We shall instead present examples where this step has been carried out by *ad hoc* methods. The procedure has been implemented on a computer, using a standard package for the numerical integration of ordinary differential equations. Among several methods, we chose the b.d.f. one. The computation of the gradient is done from explicit division formulas, which we do not derive here due to limited space, and can be carried out using the Euclidean algorithm since g is a polynomial in practice. This also implies that Ψ_n exists and is smooth on $\mathbf{R}^n$. All along the integration, a control is made on the points of the path to verify that they lie in Δ_n by computing the roots of each corresponding polynomial. If we go outside Δ_n, we use dichotomy on the stepsize, to determine accurately the crossing point on the boundary. This gives the initial condition of a lower order integration as described in section 4.

Several sets of data have been treated, most of them obtained from experimental measurements of an aircraft's high-frequency modes. Figures 4 and 5 show an example selected as being difficult to approximate, where f is described through an estimation of its first 200 Fourier coefficients. Since the maximum error occurs when $|z| = 1$, we choose to represent the functions by the 2-dimensional plots of their values around the unit disk. The function f (continuous line) and its best approximant (dotted line) of order 7 are plotted on figure 4. The computation time on a SUN station 4/110 was 13 minutes. The result is not completely satisfactory and one has to go up to order 10 to get a better approximation. This time the computation time was 17 minutes. Figure 5 shows the corresponding approximation.

7 Open questions and conclusion.

At this point, it is only fair to say that the procedure described above does not quite answer the original question, since it only ensures that we meet a *local* minimum, and not necessarily a *global* one. This deficiency is puzzling for there may be lots of local minima. Figure 6 shows already 3 local minima, when n is only equal to 2 and f is the simple rational function of order 3: $1/4\ z^{-1} + z^{-3}$.

Further investigations on this problem may be envisaged from two different viewpoints.

On one hand, it is possible to consider this 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 of the compact set Δ_n to reach all local minima and compare between them. But the precise meaning of the word "enough" depends, of course, on f and n, and we are not able to give an *a priori* bound for it. Consequently, more efficient strategies should be investigated. For instance, we can restrict ourselves to initial points lying on $\partial\Delta_n$ provided n is large enough (BARATCHART *et al*, 1990a). In examples we have met so far, initiating the algorithm from local minima on $\partial\Delta_n$ is in fact enough to exhaust the set of local minima in Δ_n. Since local minima on $\partial\Delta_n$ are solutions of the corresponding problem in Δ_{n-1} and Δ_{n-2}, thanks to lemma 1, this allows one to proceed recursively. We do not know, however, whether this property holds in some generality.

On another hand, the fact that there are several local minima may cause discrepancy in identification. For instance, there are situations (RUCKEBUSH, 78) when two distinct rational functions of order n, say r_1 and r_2, are both best approximants to f. Though these situations are exceptional (BARATCHART, 87), the L^2 identification problem at order n is not well-posed in the neighborhood of such an f, because jiggling it slightly yields a best approximant which is close to r_1 or r_2 alternatively. Whether such a phenomenon is due to f or rather a consequence of inappropriate a value for n is not yet clear. This suggests that the physical meaning of the difficulties explained above should be further analysed. In particular, it would be of interest to derive conditions on f ensuring there are no local minima except the global one, at least for n sufficiently large. A subclass of Stieljes functions, for instance, has been shown recently to have this property (BARATCHART *et al*,1990b), but a lot of work remains to be done in this direction.

A related problem is the behaviour of Ψ_n as $n \to \infty$. For instance, it is possible to prove (BARATCHART *et al*, 1990a) that all critical points of Ψ_n converge to f in H_2^- as $n \to \infty$. But the rate of convergence is likely to depend on the nature of the points (saddles or minima), and such questions are wide open.

Finally, in order to apply in a meaningful way to system theory, this technique has to be extended to the multi-input multi-output case, a question which is not yet settled. The main difficulty is to find an analogue to Ψ_n. This generalization is currently under investigation.

As a conclusion, let us express our point of view that rational approximation has something to offer in system theory and that differential tools are useful in smooth situations like the one arising here. The above algorithm is intended to be a modest contribution to this range of ideas.

Bibliography.

BARATCHART L. (1986) *Existence and generic properties of L^2 approximants for linear systems*, I.M.A. Jour. of Math. Control & Information, vol. 3, pp. 89- -101.

BARATCHART L. (1987) *Recent and New Results in Rational L^2 Approximation*, in Nato asi Series vol.34, “Modelling, Robustness and Sensitivity

Reduction in Control Systems", R.F. Curtain Ed., Springer-Verlag, 1987.

BARATCHART L., OLIVI M. (1988) *Index of critical points in rational L^2 approximation*, System & Control Letters 10.

BARATCHART L. (1989) *Interpolation and Fourier coefficients in the Hardy space H_2,* to appear in the Proc. of the MTNS, Amsterdam, June 89.

BARATCHART L., OLIVI M., WIELONSKY F. (1990a) *Asymptotic properties of critical points in rational L^2 approximation* to appear in the 9^{th} conference on Analysis and optimization of systems, Antibes, France.

BARATCHART L., OLIVI M., WIELONSKY F. (1990b) Unpublished notes.

DOOB J.L. (1953) *Stochastic Processes*, Wiley, New York.

DUREN P. (1970) *Theory of H^p Spaces*, Academic Press, New York.

FUHRMAN P. A. (1981) *Linear systems and operators in Hilbert space,* McGraw-Hill.

RUCKEBUSCH G. (1978) *Sur l'approximation rationnelle des filtres*, Rapport No 35 CMA Ecole Polytechnique.

WALSH J. L. (1962) *Interpolation and approximation by rational functions in the complex domain*, A.M.S. Publications, vol.10.