Journal of Computational and Applied Mathematics 19 (1987) 23–38
North-Holland

Toeplitz matrix techniques and convergence of complex weight Padé approximants

Alphonse P. MAGNUS
Université Catholique de Louvain, Institut Mathématique, Chemin du Cyclotron 2, B-1348 Louvain-la-Neuve, Belgium

Received 20 November 1985

Abstract: One considers diagonal Padé approximants about ∞ of functions of the form

$$f(z)=\int_{-1}^{1}(z-x)^{-1}w(x)\,dx,\quad z\notin[-1,1],$$

where w is an integrable, possibly complex-valued, function defined on $[-1, 1]$.

Convergence of the sequence of diagonal Padé approximants towards f is established under the condition that there exists a weight ω, positive almost everywhere on $[-1, 1]$, such that

$$g(x)=w(x)/\omega(x)$$

is continuous and not vanishing on $[-1, 1]$.

The rate of decrease of the error is also described.

The proof proceeds by establishing the link between the Padé denominators and the orthogonal polynomials related to ω, in terms of the Toeplitz matrix of symbol $g(\cos\theta)$.

Keywords: Padé approximation, Toeplitz matrix, orthogonal polynomials.

1. Introduction

The diagonal nth Padé approximant about ∞ of the series

$$f(z)=\sum_{k=0}^{\infty}c_k z^{-k-1} \tag{1}$$

is a rational function Q_n/P_n, with the degrees of Q_n and P_n being not greater than n, $P_n\neq 0$, and

$$P_n(z)f(z)-Q_n(z)=O(z^{-n-1})\quad\text{when } z\to\infty. \tag{2}$$

When one expresses that condition with representations $Q_n(z)=\sum_{k=0}^{n}q_k z^{n-k}$ and $P_n(z)=\sum_{k=0}^{n}p_k z^{n-k}$, one finds immediately $q_0=0$, $q_{m+1}=\sum_{k=0}^{m}c_k p_{m-k}$, $m=0, 1,\ldots, n-1$, giving Q_n when P_n is known, and

$$\sum_{k=0}^{n}c_{m-n+k}p_{n-k}=0,\quad m=n,\ n+1,\ldots,2n-1, \tag{3}$$

the typical Padé set of linear homogeneous equations for the coefficients of P_n.

This is slightly at variance with most Padé approximants definitions (as the classic [5]), especially because Q_n has not a chance to reach the degree n, and could even have a degree less than $n-1$, as there is no condition that prevents the first c_k's in (1) from vanishing. The reason for the (still formal) writings with negative powers (1) and (2) is that the Padé denominators will be linked to orthogonal polynomials without annoying changes of variable (such as $x^n P(1/x)$, etc.). However, Padé approximants are basically functions of sequences of coefficients, and their properties are still valid here (we are dealing precisely with the $[n-1/n]$ Padé approximant of $\{c_0, c_1, \ldots\}$). In particular, the function Q_n/P_n (properly defined at the zeros of P_n) is unique, independently of the rank of the matrix of (3).

We will return very soon to a thorough discussion of sets of linear equations equivalent to (3), but let us introduce first a special class of functions to be investigated.

It is extremely difficult to establish something on the convergence of sequences of diagonal Padé approximants of a function f under general conditions like "let f be holomorphic [or meromorphic] in some region", even if one accepts weak versions of convergence, as convergence of subsequences, or convergence in capacity (or both). Lubinsky set himself the formidable task of proving or disproving famous conjectures built in the 1960s, when Padé approximants became heavily used in physical calculations (see [2, §6.7]). Cf. [20,21] for a recent review and further advances.

It seems that progress can be made when f is defined in the whole complex plane, up to a finite number of exceptional points: classes of entire functions, functions with a finite number of some kind of essential singularities, functions with a finite number of branch-points (in which case the appropriate domain of f is a Riemann surface [30]), or even functions analytic outside a set of vanishing capacity [36,37,39].

Other manageable classes of functions are defined through strong monotonicity properties of their sequence of coefficients $\{c_k\}$: this includes the Stieltjes class, the Schoenberg class, and... that's all! (see [2, Chap. 5; 12]).

One can then take one of these successful classes of functions and try to extend them by more or less infinitesimal steps.

This will be done here, starting from Stieltjes (or Markov) class

$$f(z) = \int_E (z-t)^{-1} \, \mathrm{d}\alpha(t), \quad z \notin E$$

where α is a bounded increasing function of real support E.

The importance of such functions is well known, they may be generated from spectral investigations of self-adjoint operators H by $f(z) = (\phi_0, (z-H)^{-1}\phi_0)$ (see [8,11,14]). They are also related (up to changes of variables) to Herglotz and Nevanlinna classes of functions. If we restrict ourselves to diagonal Padé approximants, convergence is ensured if E is bounded (if not, one must study the determinateness of the moments problem and related questions [19, p. 310]). One can then always suppose that E is in $[-1, 1]$.

In what follows, the form under investigation will be

$$f(z) = \int_{-1}^{1} (z-x)^{-1} w(x) \, \mathrm{d}x, \quad z \notin [-1, 1], \tag{4}$$

where w is allowed (with many restrictions that will be explained) to take *complex* values.

These forms may appear when one investigates functions analytic on $\bar{\mathbb{C}}\setminus[-1, 1]$ with $f(\infty)=0$: such a function can always be written

$$f(z)=(2\pi\mathrm{i})^{-1}\int_C(z-t)^{-1}f(t)\,\mathrm{d}t, \quad z\notin[-1, 1]$$

from Cauchy formula, where C encloses $[-1, 1]$ but not z. Now, if f is regular enough when z approaches $[-1, 1]$, C may shrink to upper and lower sides of $[-1, 1]$, giving

$$f(z)=\int_{-1}^{1}(z-t)^{-1}(f_-(t)-f_+(t))/(2\pi\mathrm{i})\,\mathrm{d}t,$$

where $f_+(t)$ and $f_-(t)$ are the limits of $f(t\pm\mathrm{i}\epsilon)$ when $\epsilon\to 0$ (see [12, p. 247–248; 8, Chap. 1]).

Now, if one wants to approximate $f(z)=(\phi_0, (z-H)^{-1}\phi_0)$ when H is not a self-adjoint operator, as in relaxation or dissipative problems [11,14] where $H=H_1-\mathrm{i}H_2$ with H_1 and H_2 self-adjoint and H_2 semipositive definite, complex representations may also be expected. It is easy to show that such an $f(z)$ satisfies $\operatorname{Im} f(z)<0$ when $\operatorname{Im} z>0$ (solve $(z-H)\,\psi=\phi_0$ in the domain of H and write $f(z)=((z-H)\psi, \psi)=(\psi, (z-H^+)\psi)$). A series (1) still holds with complex $c_k=(\phi_0, H^k\phi_0)$, but the existence of a form like (4) is still not clear (relaxation moment problem [11,14]).

Diagonal Padé approximants of functions of the form (4) have been studied by Nuttall and Wherry [31]. One of their theorems is based on properties of the function $w(t)/\omega(t)$, with $\omega(t)=(1-t^2)^{-1/2}$ and the method of proof relates the denominators P_n to the first kind Chebyshev polynomials. Very similar techniques will be used here, but with a somewhat broader choice of comparison weight functions ω.

Much more general situations are considered by Stahl [33,34], as far as the support of w is concerned (actually, $w(t)\,\mathrm{d}t$ can be extended to a measure $\mathrm{d}\mu(t)$, but this does not seem to be very useful here). From a simple case treated in [35], one learns that w should avoid 0 in some way... .

2. Matrix expression of diagonal Padé approximants

We take again the problem of the determination of the polynomials P_n and Q_n in (2). Instead of using the basis $\{x^k\}$, it will be found much more convenient to express P_n and Q_n as combinations of polynomials Π_k, $k=0, 1, 2,\ldots$, where Π_k will be a suitably chosen polynomial of exact degree k:

$$P_n(z)=\pi_{0,n}\Pi_n(z)+\cdots+\pi_{n,n}\Pi_0(z). \tag{5}$$

The expansion about ∞ of $P_n(z)f(z)$ could be obtained by returning to powers of z, which would involve directly the c_k's of (1), but it is good practice [4,7] to introduce the following functional first defined on the space of polynomials:

$$c(x^k)=c_k, \quad k=0, 1,\ldots$$

and it is not hard to guess that the values $c(\Pi_k)$ will become our building blocks. Moreover, the representation (4) allows to define c on the class of bounded functions:

$$c(\phi)=\int_{-1}^{1}\phi(t)w(t)\,\mathrm{d}t, \quad \phi\in L_\infty,$$

and that $f(z)$ itself can be expressed as

$$f(z)=c\left((z-t)^{-1}\right),$$

a writing that is only formal, should (4) not be available [4,7]. Then,

$$Q_n(z)=c\left((z-t)^{-1}(P_n(z)-P_n(t))\right),$$
$$P_n(z)f(z)-Q_n(z)=c\left((z-t)^{-1}P_n(t)\right)$$

follow easily. Expansion in negative powers of z gives

$$P_n(z)f(z)-Q_n(z)=\sum_{k=0}^{\infty}c\left(t^kP_n\right)z^{-k-1},$$

so that if we want to have (2), P_n must be orthogonal to any polynomial of degree less than n, with respect to c, that is to w. This connection between Padé denominators and general, or formal orthogonal polynomials (not restricted to a positive weight) is well known (cf. [5,4,7,...]), as well as the consequence for the error

$$\begin{aligned} f(z)-Q_n(z)/P_n(z)&=c\left((z-t)^{-1}(P_n(t))^2\right)/(P_n(z))^2\\ &=\int_{-1}^{1}(z-t)^{-1}(P_n(t))^2w(t)\,\mathrm{d}t/(P_n(z))^2. \end{aligned}\tag{6}$$

The orthogonality relations are therefore

$$c(\Pi_mP_n)=\sum_{k=0}^{n}c(\Pi_m\Pi_k)\pi_{n-k,n}=0,\quad m=0,\dots,n-1.\tag{7}$$

If we take $\Pi_k(x)=x^k$, this is exactly (3). One could also deduce (7) from (3) by equivalence transformations (multiplications by triangular matrices). The reason for choosing a general basis $\{\Pi_k\}$ is that it will be possible to get a matrix $[c(\Pi_m\Pi_k)]$ more 'readable' than $[c_{m+k}]$ with respect to approximate inversion. Also, condition number reduction can be invoked [10, Section 3 and 4]. Indeed, let us define the matrix

$$M_n=\left[c(\Pi_m\Pi_k)\right]_{m,k=0}^{n-1}.\tag{8a}$$

The discussion of the behaviour of P_n through its coefficients $\pi_{j,n}$ will necessarily involve the inversion of M_n. This task of estimating the elements of $(M_n)^{-1}$ is hopeless for Hankel matrices, hypersensitive to perturbation, as well known [12, p. 344].

On the opposite, if the polynomials Π_k's are already orthogonal with respect to a weight function ω on $[-1, 1]$, and if w is 'not too far' from ω, the matrix M_n will be 'not too far' from a diagonal matrix, and the discussion of (7) will be easy. The next sections are devoted to more accurate versions of this idea.

Final algebraic aspects include that, if det $M_n\neq 0$, P_n (called a regular orthogonal polynomial [7, p. 47]), has exact degree n and is uniquely determined up to a nonzero multiplicative factor, say $\pi_{0,n}$; and that, if both det M_n and det $M_{n+1}\neq 0$, P_n is not orthogonal to itself:

$$h_n=c\left(P_n^2\right)=\pi_{0,n}c(\Pi_nP_n)=(\pi_{0,n})^2\det M_{n+1}/\det M_n\neq 0.$$

The approximant itself has a very elegant expression, due to Nuttall (see [4, p. 17] with $v\equiv P_n$)

$$Q_n(z)/P_n(z)=\mu^{\mathrm{T}}(zM_n-N_n)^{-1}\mu\tag{9}$$

where μ^{T} (the transposed of μ) is $[c(\Pi_0),\dots,c(\Pi_{n-1})]$, and where

$$N_n = \left[c\left(t\Pi_m(t)\Pi_k(t)\right)\right]_{m,k=0}^{n-1}. \tag{8b}$$

3. Infinite matrices as operators on l^2

Convergence is linked to some concept of limit which would allow to conclude that $(zM_n - N_n)^{-1} \to (zM - N)^{-1}$ when $n \to \infty$, where M and N are the infinite matrices

$$M = \left[c(\Pi_m\Pi_k)\right]_{m,k=0}^{\infty}, \qquad N = \left[c\left(t\Pi_m(t)\Pi_k(t)\right)\right]_{m,k=0}^{\infty}. \tag{10}$$

This asks for a preliminary careful definition of algebraic operations on infinite matrices.

Theorem 3.1. *A bounded linear operator on the space l^2 of complex square summable sequences $x = \{x_k\}_{k=1}^{\infty}$, $\|x\|^2 = \sum_{k=1}^{\infty}|x_k|^2 < \infty$, can always be represented by an array $\{a_{m,k}\}_{m,k=1}^{\infty}$ of complex numbers in such a way that*

$$Ax = \left\{\sum_{k=1}^{\infty} a_{m,k}x_k\right\}_{m=1}^{\infty}, \qquad \forall x \in l^2$$

with $\|A\| = \sup_{\|x\|\leqslant 1}\|Ax\| < \infty$; *this representation is unique.*

Theorem 3.2. *The algebra of the infinite matrix representations of bounded linear operators on l^2 is the natural extension of the usual matrix algebra*:

$$A + B = \{a_{m,k} + b_{m,k}\}_{m,k=1}^{\infty}, \qquad \|A+B\| \leqslant \|A\| + \|B\|;$$

$$AB = \left\{\sum_{j=1}^{\infty} a_{m,j}b_{j,k}\right\}_{m,k=1}^{\infty}, \qquad \|AB\| \leqslant \|A\|\,\|B\|.$$

The proofs can be found in many textbooks of functional analysis, sometimes as exercises. Kantorovitch and Akilov [17], Chapter 6, §1; and, Chapter 10, §1] give a detailed exposition aimed at numerical analysis applications. Of course, the subject is treated with considerable extension in specialized works on sequence spaces (see references to Cooke, McDuffee, Köthe and Toeplitz, Zeller... in [22]); the key ideas go back to Schur, Landau, F. and M. Riesz, Hilbert, Hellinger and Toeplitz in the early 1910s.

Definition 3.3. A Toeplitz matrix in Grenander and Szegö's sense is an infinite matrix of the form

$$T_{\mathrm{GS}}(f, \{\phi_k\}, \mathrm{d}\mu) = \left\{\int_S f(t)\overline{\phi_m(t)}\phi_k(t)\,\mathrm{d}\mu(t)\right\}_{k,m=1}^{\infty}$$

where μ is a positive measure on S, $\{\phi_k\}$ is a complete orthonormal sequence of $L_2(\mathrm{d}\mu)$, and f (the symbol of T) is a bounded function.

Theorem 3.4. *For fixed* $\{\phi_k\}$ *and* $\mathrm{d}\mu$, *the algebra of Toeplitz matrices in Grenander and Szegö's sense is the algebra of their symbols*:

$$T_{\mathrm{GS}}(f+g, \{\phi_k\}, \mathrm{d}\mu) = T_{\mathrm{GS}}(f, \{\phi_k\}, \mathrm{d}\mu) + T_{\mathrm{GS}}(g, \{\phi_k\}, \mathrm{d}\mu),$$
$$T_{\mathrm{GS}}(fg, \{\phi_k\}, \mathrm{d}\mu) = T_{\mathrm{GS}}(f, \{\phi_k\}, \mathrm{d}\mu) T_{\mathrm{GS}}(g, \{\phi_k\}, \mathrm{d}\mu).$$

Moreover, $\| T_{\mathrm{GS}}(f, \{\phi_k\}, \mathrm{d}\mu) \| = \| f \|_\infty$.

The proof is in Grenander and Szegö's book [13, Chapter 8]. It is essentially based on Parseval relations. A useful by-product is an interpretation of Tx: associate to the sequence $x \in l^2$ the (generalized) Fourier series $\chi(t) = \Sigma_{k=0}^{\infty} x_k \phi_k(t)$. Then Tx is nothing else than the sequence of the Fourier coefficients of $f\chi$:

$$\sum_{k=1}^{\infty} (T_{\mathrm{GS}}(f, \{\phi_k\}, \mathrm{d}\mu)_{m,k} \int_S \overline{\phi_k(t)} \chi(t)\, \mathrm{d}\mu(t) = \int_S \overline{\phi_m(t)} f(t) \chi(t)\, \mathrm{d}\mu(t),$$
$$m = 1, 2, \ldots .$$

Now, here is why this special class of matrices is useful in Padé approximation:

Remark 3.5. When $c(\phi) = \int_{-1}^{1} \phi(t) w(t)\, \mathrm{d}t$ and $\{\Pi_k\}$ is the sequence of the orthonormal polynomials related to a nonnegative function ω, the matrices M and N of (10) are Toeplitz matrices in Grenander and Szegö's sense of symbols $g(t) = w(t)/\omega(t)$ and $\mathrm{t}g(t)$. In particular, if $0 < C_1 < |g(t)| < C_2 < \infty$ on $-1 \leqslant t \leqslant 1$, $zM - N$ is invertible when $z \notin [-1, 1]$, and one has

$$f(z) = \mu^{\mathrm{T}} (zM - N)^{-1} \mu, \quad z \notin [-1, 1]. \tag{11}$$

It may seem that a big step has been achieved, as the matrix formula (9) has a meaning when $n = \infty$. But nothing more has been done: (9) gives a matrix form of the nth diagonal Padé approximant of f, (11) is the matrix form of f itself. Convergence of the matrix forms is no more obvious than convergence of the functional forms. The issue is that the finite matrices of (8a, b) are obtained from the infinite matrices of (10) after multiplication by projectors

$$E_n = [u_k \delta_{k,m}]_{k,m=1}^{\infty}, \tag{12}$$

with $u_1 = u_2 = \cdots = u_n = 1$ and $u_k = 0$ for $k > n$:

$$zM_n - N_n = E_n (zM - N) E_n.$$

In general, even if an infinite matrix A is invertible, it does not follow that A^{-1} is the limit of the inverses $(E_n A E_n)^{-1}$ when $n \to \infty$. An infinite number of these inverses could even not exist: think of an invertible skew symmetric infinite matrix, for instance.

4. Convergence of approximate inverses

This section is devoted to conditions ensuring $(A_n)^{-1} \to A^{-1}$, as this will be translated in convergence of Padé approximants.

First, here are some precisions on the tools that will be used:

Definitions 4.1. (i) Weak convergence of sequences of $l^2 x^{(n)}$ towards $x \in l^2$ holds if $y^{\mathrm{T}} x^{(n)} = \Sigma_{k=1}^{\infty} y_k x_k^{(n)} \to y^{\mathrm{T}} x$ when $n \to \infty$, for any $y \in l^2$.

(ii) Strong or ordinary convergence of sequences of $l^2 x^{(n)} \to x \in l^2$ holds if $\| x - x^{(n)} \| \to 0$ when $n \to \infty$.

(iii) Strong, or ordinary convergence of bounded linear operators on $l^2 A_n \to A$ holds if $A_n x \to Ax$ for any $x \in l^2$.

(iv) Norm convergence of bounded linear operators on l^2 holds if $\| A - A_n \| \to 0$ when $n \to \infty$.

Remarks 4.2. (i) The first definition uses the fact that bounded linear forms on l^2 can always be represented by a sequence $y \in l^2$ in such a way that the value of the form at x is $y^{\mathrm{T}} x = \sum_{k=1}^{\infty} y_k x_k$. There is of course a more popular representation of forms on l^2 by sequences z such that the value of the form at x is the scalar product $(z, x) = \sum_{k=1}^{\infty} \bar{z}_k x_k$ but complex conjugates appear nowhere in Padé constructions. This is the basic reason why the present study is restricted to *real* Toeplitz matrices in Grenander and Szegö's sense, and therefore to integral forms (4) limited to real sets of integration. An extension to complex sets of integration should be most useful [30]; see also López [19, p. 312] encountering the same difficulties.

(ii) As remarked by Widom [41, p. 338], the 'strong', or ordinary convergence of operators is the counterpart of the weak convergence of sequences. Everything would be easy with norm convergence but this occurs practically only with compact operators, as it will be shown. The reason why norm convergence will normally not be allowed is that the sequence of projectors E_n (12) tends only in the ordinary sense towards the unit matrix $I = [\delta_{m,k}]$ and this implies that the limits $M_n \to M$ and $N_n \to N$ are usually not reached in the norm sense. However, it can be said that the norms of the bounded operators A_n are uniformly bounded: $\| A_n \| \leqslant C$, if $A_n \to$ the bounded operator A (Banach Steinhaus theorem; see [17 Ch. 10, §1, Theorem 2] for a typical application).

The problem is therefore to ensure $(A_n)^{-1} \to A^{-1}$ when $A_n \to A$ in the ordinary sense. The answer is given by Kantorovitch's theorem [16; 17, Ch. 14] given here in its generality.

Theorem 4.1. Let A be a bounded linear operator of a normed linear space X *onto* a normed linear space Y. Consider a sequence of bounded invertible operators $\{A_n\}$ of X onto Y such that $A_n \to A$ (consistency), with uniformly bounded inverses: $\|(A_n)^{-1}\| \leqslant C$ (stability).

Then, A^{-1} exists and $(A_n)^{-1} \to A^{-1}$ when $n \to \infty$ (convergence).

Proof. The proof is so simple that it is recalled here: one must show that, for any $y \in Y$, the sequence $\{(A_n)^{-1} y\}$ has a limit in X. As A acts onto Y, there is at least one $x \in X$ such that $Ax = y$. Let us study the sequence $\{(A_n)^{-1} y - x\}$. One has $(A_n)^{-1} y - x = (A_n)^{-1}[y - A_n x] = (A_n)^{-1}[Ax - A_n x]$. Now, by the consistency condition, the norm of the second factor converges towards zero, so that x is indeed the (therefore unique) limit of $(A_n)^{-1} y$, thanks to the stability condition. □

If X and Y are Banach spaces, it is not necessary to check that A acts onto Y: this a consequence of the stability condition and completeness of X. The theorem is then summarized by the well known sentence: "if consistency holds, stability is equivalent to convergence" ($(A_n)^{-1} \to A^{-1}$ implies $\|(A_n)^{-1}\| \leqslant C$, from Banach–Steinhaus theorem).

It will be necessary to change some details of the theorem, because the range of the matrices M_n and N_n is not the whole space l^2, but only a part of it, and even a finite-dimensional subspace in the cases of interest. We will therefore consider the sequence of approximations $A_n = E_n A E_n$, where E_n is a linear projector on l^2 (i.e., $(E_n)^2 = E_n$). With the special choice (12), $E_n A E_n$ appears as an infinite matrix filled with zeros with the exception of the n, n upper left block:

$$E_n A E_n = \begin{bmatrix} a_{1,1} & \cdots & a_{1,n} & 0 & 0 & \cdots \\ \vdots & & \vdots & \vdots & \vdots & \\ a_{n,1} & \cdots & a_{n,n} & 0 & 0 & \cdots \\ 0 & \cdots & 0 & 0 & 0 & \cdots \end{bmatrix}.$$

If $E_n A E_n$ is one-ton-one on $E_n l^2$, it is natural to note $(E_n A E_n)^{-1}$ the operator reverting the connection. This last operator is of course only defined on $E_n l^2$. If E_n is self-adjoint (orthogonal projector), the connection with the Moore–Penrose pseudoinverse is of interest (cf. [25, eqs. (30)]).

Theorem 4.2. *Let A be a bounded invertible operator on l^2. Let $\{E_n\}$ be a sequence of linear projectors tending to the identity operator: $E_n \to I$ when $n \to \infty$. If $\|(E_n A E_n)^{-1}\| \leqslant C$ when n is large enough, then*

$$(E_n A E_n)^{-1} E_n \to A^{-1} \quad \text{when } n \to \infty.$$

Indeed, one just has to write, for any $x \in l^2$,

$$\begin{aligned}(E_n A E_n)^{-1} E_n x - A^{-1}x &= (E_n A E_n)^{-1}[E_n x - E_n A E_n A^{-1}x] + (E_n - I)A^{-1}x \\ &= (E_n A E_n)^{-1} E_n [A - A E_n] A^{-1}x + (E_n - I)A^{-1}x\end{aligned}$$

and to remark that each term tends towards 0.

The theorem is also given by Widom [41, Theorem 7.1].

Now, if A has all the nice properties, we try to find neighbouring matrices (operators) sharing them. The important subclass of bounded compact operators (or completely continuous operators) will be found very useful. Instead of giving a definition followed by numerous consequences, only the properties which will be used later on are stated here:

Property 4.3. *A compact operator on l^2 transforms any weakly convergent sequence of l^2 into a strongly convergent one; multiplication by a compact operator transforms ordinary convergence of operators into norm convergence: if K is compact, then*
- *if $y^{\mathrm{T}}x^{(n)} \to y^{\mathrm{T}}x$ for any $y \in l^2$, then $\| Kx^{(n)} - Kx \| \to 0$,*
- *if $A_n \to A$ bounded on l^2, then $\| KA_n - KA \|$ and $\| A_n K - AK \| \to 0$.*

The derivation is basically in Kantorvitch and Akilov [17, Chapter 10, §1]. Widom's [41, §8] starting point is to define the compact operators as the closure of the finite rank operators with respect to the norm topology. One can indeed recognize a compact operator K by

$$\| K - E_n K \|, \quad \| K - K E_n \| \quad \text{and} \quad \| K - E_n K E_n \| \to 0 \quad \text{when } n \to \infty$$

where E_n is the finite rank projector (12) (also in [17, Chapter 10, §1, Theorem 1]). For instance, a band matrix whose elements tend to zero

$$K_{m,k} = 0 \quad \text{if } |k-m| > d; \qquad K_{m,k} \to 0 \quad \text{when } m \text{ and } k \to \infty, \quad |k-m| \leqslant d$$

represents a compact operator.

Theorem 4.4. *Let A be a bounded operator on l^2 and $\{E_n\}$ a sequence of projectors tending towards the identity operator. Moreover, let us suppose that $\|(E_n A E_n)^{-1}\| \leqslant C$ for n large enough (so that A is invertible and $(E_n A E_n)^{-1} \to A^{-1}$).*

Then, if K is compact, $\|(E_n(A+K)E_n)^{-1}\| \leqslant C'$ for n large enough if and only if $A+K$ is invertible.

The new information given by this theorem is a class of operators for which there is no spuriously singular behaviour of the approximate inverses: if B^{-1} is bounded ($B = A + K$), then the stability property will hold for the $(E_n B E_n)^{-1}$'s, whereas, up to now, this stability property had to be put in the hypotheses.

Proof. as just remarked, the new feature to be proved is that invertibility of $A+K$ implies the uniform boundedness of the approximate inverses $(E_n(A+K)E_n)^{-1}$. Let us show that, for any $y \in E_n l^2$, it is possible to solve $E_n(A+K)E_n x = y$, provided n is large enough, and that a relation $\|x\| \leqslant C' \|y\|$ holds, with C' independent of n. Indeed, the equation for x is also $(I + (E_n A E_n)^{-1} E_n K E_n)x = (E_n A E_n)^{-1} y$. From the Properties 4.3 of the compact operator K, this last equation can be written $(I + A^{-1}K + \Delta_n)x = z^{(n)}$, where $\|\Delta_n\| \to 0$ when $n \to \infty$. The equation can therefore be solved from the value of n for which $\|\Delta_n\| < 1/\|(I + A^{-1}K)^{-1}\|$ onwards. And as $z^{(n)} \to A^{-1}y$, $\|x\| \leqslant (I + A^{-1}K)^{-1}\| \; \|A^{-1}\|(1+\epsilon)\|y\|$, with arbitrary $\epsilon > 0$, provided n is large enough. □

Our study of the approximate inverses needed in (9) will indeed exhibit $zM - N$ as a 'nice' matrix augmented by a compact matrix. Toeplitz matrices enter now the stage.

5. Relation with Toeplitz matrices

Let us return to the matrices M and N of (10) with the hypothese that the polynomials Π_m are orthonormal with respect to a weight function ω on $[-1, 1]$. As well known, such polynomials satisfy a three-term recurrence relation

$$\alpha_m \Pi_{m-1}(x) - (x - \beta_m)\Pi_m(x) + \alpha_{m+1}\Pi_{m+1}(x) = 0 \tag{13}$$

where the β_m's are real and the α_m's are positive.

Lemma 5.1. *If the coefficients α_m and β_m of* (13) *satisfy*

$$\alpha_m \to \tfrac{1}{2}, \quad \beta_m \to 0 \quad \text{when } m \to \infty \tag{14}$$

and if $w/\omega = p$ is a polynomial of degree d, then M and N are band matrices whose elements have limits in such a way that

$$\sum_{k=m-d}^{m+d} M_{m,k}\, \mathrm{e}^{\mathrm{i}(k-m)\theta} \to p(\cos\theta),$$

$$\sum_{k=m-1-d}^{m+d+1} N_{m,k}\, \mathrm{e}^{\mathrm{i}(k-m)\theta} \to \cos\theta\, p(\cos\theta)$$

when $m \to \infty$.

The proof is elementary [22], using the facts that M and N are Toeplitz matrices in Grenander and Szegö's sense of symbols $p(t)$ and $tp(t)$ (from Remark 3.5), that the T_{GS} matrix of symbol t is the tridiagonal matrix of the α's and the β's, and that M and N are polynomial functions of this matrix.

Definition 5.2. *An ordinary Toeplitz matrix of symbol* $f \in L_\infty[-\pi, \pi]$ *has the form*

$$T(f) = [a_{m-k}]_{m,k=1}^{\infty},$$

where $f(\theta) = \sum_{-\infty}^{\infty} a_k\, e^{ik\theta}$ *(at least formally).*

Remark 5.3. Under the conditions of Lemma 5.1, $xM - N = T((x - \cos\theta)p(\cos\theta)) + K$, where K is compact.

Of course, one progresses slowly towards an application of Theorem 4.4, but ordinary Toeplitz matrices must not be confused with the following definition.

Definition 5.4. The doubly infinite Toeplitz matrix of symbol $f \in L_\infty[-\pi, \pi]$ has the form

$$\tilde{T}(f) = [a_{m-k}]_{m,k=-\infty}^{\infty},$$

where $f(\theta) = \sum_{-\infty}^{\infty} a_k\, e^{ik\theta}$.

There is nothing fundamental in using indexes going from $-\infty$ to ∞, one could use odd and even indexes (and sines and cosines in the Fourier expansion of f). As $\{(2\pi)^{-1}\exp(ik\theta)\}$, $-\infty < k < \infty$, is a complete orthonormal set of $L_2[-\pi, \pi]$, $\tilde{T}$ is a Toeplitz matrix in Grenander and Szegö's sense, so that it is invertible only if $f^{-1} \in L_\infty[-\pi, \pi]$.

Things are not so simple with 'ordinary' Toeplitz matrices, where one must consider the following theorem.

Theorem 5.5. *Let* f *be continuous and periodic (i.e.,* $f(-\pi) = f(\pi)$*). The ordinary Toeplitz matrix* $T(f)$ *is then invertible only if* $\log f$ *is also continuous and periodic.*

Then, with the two expansions

$$1/f(\theta) = \rho_1(e^{i\theta})\rho_2(e^{-i\theta}) = \left(\sum_0^\infty \xi_k\, e^{ik\theta}\right)\left(\sum_0^\infty \eta_k\, e^{-ik\theta}\right)$$

(Wiener–Hopf factorization), obtained by exponentiating separately the parts with positive and negative powers of $\exp(i\theta)$ *in the Fourier series of* $-\log f$*, one has the triangular factorization of* T^{-1}:

$$(T(f))^{-1}_{k,m} = \sum_{j=1}^{\min(k,m)} \xi_{k-j}\eta_{m-j}.$$

formally (element-wise), as the triangular factors could be unbounded.

This is only a special result in Toeplitz operators theory, that one can study in Douglas [6, Chapter 7, theorem 7.26 and 7.27] or Widom [40, Theorem 5]. Remark that the Wiener–Hopf factors are bounded if the Fourier series of $\log f$ is absolutely convergent, but the theorem holds without this condition. Bourgain [3] has shown a way to construct bounded factors but they could contain Blaschke products and have zeros in the unit disk, whereas ρ_1 and ρ_2 must be analytic without zero in the unit disk. The periodicity condition for log f can be translated by $W=0$, where W is the winding number about the origin of the curve $\{f(\theta)\}$, $-\pi \leqslant \theta \leqslant \pi$.

Remark 5.6. If $a_{-k}=a_k$, $k \geqslant 1$, the conditions of Theorem 5.5 are fulfilled if f is continuous and does never vanish on $[-\pi, \pi]$ (see the remark in [40, p. 198]).

And at last, here is a theorem giving conditions for $T(f)$ to be 'a nice' matrix.

Theorem 5.7. *Under the conditions of Theorem* 5.5 *on* f, *one has*

$$\|(E_nT(f)E_n)^{-1}\| \leqslant C, \quad n \text{ large enough},$$

where (E_n) *is the sequence of finite rank orthogonal projectors* (12).

The proof [1; 41, §6] makes a clever use of Theorem 4.4, where Hankel matrices play the role of compact perturbations. Here is a short sketch of Ambarcumjan's proof: write

$$\tilde{T}(f) = \begin{array}{c} \\ -\infty \\ 0 \\ 1 \\ \infty \end{array} \begin{array}{c} \begin{array}{cccc} -\infty & 0 & 1 & \infty \end{array} \\ \begin{bmatrix} T(f) & & & H_2 \\ & & & \\ & & & \\ H_1 & & & T(f) \end{bmatrix} \end{array} .$$

H_1 and H_2 are Hankel matrices that represent compact operators, thanks to the continuity condition of f (see [41, Theorem 9.1]). $\tilde{T}$ and $\tilde{T}-H_1-H_2$ are both invertible (the second one is reduced to two diagonal blocks equivalent to T), so that if one is 'nice', so will be the other. Let us work with projectors $\tilde{E}_n$, similar to (12) but keeping also elements with negative or zero index. Then, the matrices $\tilde{E}_n\tilde{T}\tilde{E}_n$ have uniformly bounded inverses, as they are all equivalent to T (acting on $[x_n, x_{n-1}, \ldots, x_1, x_0, x_{-1}, \ldots]^{\mathrm{T}}$)! Therefore, the same holds when n is large enough for $\tilde{E}_n(\tilde{T}-H_1-H_2)\tilde{E}_n$ and for its finite diagonal block $E_nTE_n \ldots$.

6. A convergence theorem for diagonal Padé approximants

Here is an application of the preceding theory to complex weight Padé approximants:

Theorem 6.1. *Let f be the function*

$$f(z) = \int_{-1}^{1} (z-t)^{-1} w(t)\,\mathrm{d}t, \quad z \notin [-1, 1],$$

where w is an integrable, possibly complex-valued, function. Suppose that there exists a real function ω, integrable and positive almost everywhere on $[-1, 1]$, *such that*

$$g = w/\omega$$

is continuous and non vanishing on $[-1, 1]$.

Then, the diagonal Padé approximants Q_n/P_n *about* ∞ *have the following properties*:

(i) *When n is large enough,* P_n *has exact degree n and is uniquely determined, up to a multiplicative constant. Moreover,*

$$h_n = \int_{-1}^{1} (P_n(t))^2 w(t)\, \mathrm{d}t \neq 0.$$

(ii) *The sequence of the diagonal Padé approximants converges towards f uniformly in any compact of* $\overline{\mathbb{C}} \setminus [-1, 1]$.

The following points are also of interest:

(iii) *With an appropriate choice of the multiplicative constant in* P_n, $h_n = 1$,

$$P_n(z)/\Pi_n(z) \xrightarrow[n\to\infty]{} \rho\left(z - (z^2-1)^{1/2}\right)$$

uniformly in any compact of $\overline{\mathbb{C}}/[-1, 1]$, *where the* Π_n*'s are the orthonormal polynomials related to* ω, *and where* ρ *is the analytic function without zero in the unit disk, satisfying formally* $\rho(\exp(\mathrm{i}\theta))\rho(\exp(-\mathrm{i}\theta)) = 1/g(\cos\theta)$. *Moreover,*

$$\int_{-1}^{1} |P_n(t)|^2 |w(t)|\, \mathrm{d}t \leqslant C < \infty. \tag{15}$$

(iv) *When n is large enough, a three-terms recurrence relation*

$$a_{n+1}P_{n+1}(x) = (x - b_n)P_n(x) - a_n P_{n-1}(x) \tag{16}$$

holds, and one has

$$a_n \to \tfrac{1}{2}, \quad b_n \to 0 \quad \textit{when } n \to \infty. \tag{17}$$

(v) *One can bound the asymptotic rate of convergence by*

$$\limsup_{n\to\infty} |f(z) - Q_n(z)/P_n(z)|^{1/n} \leqslant |z - (z^2-1)^{1/2}|^2, \quad z \notin [-1, 1].$$

Remarks 6.2. (i) A special case of the theorem has been published in [22]. Remark that the arthitectural design of the theorem is as in Nuttall and Wherry [31, Section 2].

(ii) The requirement that g is continuous and does not vanish on $[-1, 1]$ is strong, as it means that $|g|$ is bounded from above and from below: $0 < C_1 < |g(x)| < C_2 < \infty$. In the real nonnegative case $g(x) \geqslant 0$, the theorem is still valid when $g(\cos\theta)$ and $1/g(\cos\theta)$ are allowed to vanish like a (nonnegative) trigonometric polynomial at a finite number of points [24,28]. It seems here that something can be done if g vanishes or becomes infinite in such a way that the winding number of the curve $\{g(\cos\theta)\}$ is still well defined (cusp or asymptote) in order to avoid the phenomenon of [35]. However, the techniques of proof of the present paper are still not powerful enough to handle these cases. In particular, all the "for n large enough" that permeate the theorems should be replaced by more accurate statements but see [42, Ch. 4].

In the real nonnegative case, continuity of g is not essential in [24; 27, §6.1, Theorem 27; 28] (Riemann integrability of g suffices). This is perhaps also a weakness of the methods of proof used here (think that complex Jacobi weights are not covered by Theorem 6.1, see also the remarks by Nuttall and Wherry [31, Section 4]. However, if $g \geqslant 0$ satisfies a Lipschitz condition on a part Δ of $[-1, 1]$, it is possible to find asymptotic estimates of $P_n(x)$ for x in Δ [24, III]. According to Paul Nevai [29], similar estimates could be found here.

(iii) Point (i) in Theorem 6.1 tells that all the P_n's are regular formal orthogonal polynomials [7, p. 47] provided n is large enough, and this implies the recurrence relation (16). For small values of n, some P_n's may fail to be regular, but a three-term recurrence relation linking regular P_n's still exists [7, p. 71] (see also [30, Lemma 7.1 and 7.2]), so that the diagonal Padé approximants can always be obtained as approximants of a continued fraction. The property (17) allows then the use of *continued fraction modification* techniques [15,38] accelerating the convergence. Other continued fraction aspects are presented in [11,14,18].

(iv) The property (17) will be shown to follow from (14) which is the consequence of a recent deep result [32]:

Rahmanov's Theorem. *If ω is integrable and positive almost everywhere on* $[-1, 1]$, *then among other results, the related orthonormal polynomials Π_n's are linked by a recurrence relation* (13) *whose coefficients α_n and β_n satisfy* (14).

This theorem, which represents a major advance in orthogonal polynomials theory, has been commented, simplified, extended in [23,28], among other works. It is also used in Padé theory in [19].

Proof of Theorem 6.1. In order to reach (ii), one must show that the matrix $zM - N$ of (11) has the 'nice' property

$$\| (E_n(zM-N)E_n)^{-1} \| \leqslant C \quad \text{when } n \text{ is large enough,}$$

where E_n is the projector (12) taking the n first elements of a sequence of l^2:

$$E_n x = (x_1, x_2, \dots, x_n, 0, 0, \dots),$$

remark that $\| E_n \| = 1$. More generally, this will be established for any matrix of the form

$$A = \left[\int_{-1}^{1} \phi(t) \Pi_m(t) \Pi_k j(t) w(t) \, \mathrm{d}t \right]_{m,k=0}^{\infty}$$

where ϕ is a continuous nonvanishing function on $[-1, 1]$, and where the Π_n's are the orthonormal polynomials related to ω. From the Definition 3.3 of Toeplitz matrices if Grenander and Szegö's sense, one has immediately

$$A = T_{\mathrm{GS}}(\phi g, \{\Pi_m\}, \omega(t) \, \mathrm{d}t)$$

and, from Theorem 3.4, A has a bounded inverse.

Now, we will extend the Lemma 5.1: under the conditions of the Theorem 6.1,

$$A = T(\Phi) + K$$

where $\Phi(\theta) = \phi(\cos\theta) g(\cos\theta)$ and K is *compact*. Indeed, one must show that $\| K - E_n K \| \to 0$ when $n \to \infty$ (see Property 4.3), K being $A - T(\Phi)$. For any $\epsilon > 0$, there is a polynomial p such that $\| \Phi - p \|_\infty \leqslant$ say $\frac{1}{5}\epsilon$, as Φ is continuous on $[-1, 1]$. Also, $\| B - A \| \leqslant \frac{1}{5}\epsilon$, where $B = T_{\mathrm{GS}}(p)$ is a band matrix. Now, thanks to Rahmanov's Theorem (see Remark 6.2 (iv)), we are in a position to apply Lemma 5.1: $B = T(p(\cos\theta)) + J$, from Remark 5.3, where J is compact. Therefore (Property 4.3), there is a N such that, if $n > N$, $\| J - E_n J \| \leqslant \frac{1}{5}\epsilon$, and we just have to write

$$K - E_n K = A - B + T(p) - T(\Phi) - E_n(A - B) + J - E_n J - E_n(T(p) - T(\Phi))$$

to ensure $\| K - E_n K \| \leqslant \epsilon$. As A and $T(\Phi)$ are invertible, $\|(E_n A E_n)^{-1}\| \leqslant C < \infty$ for n large enough follows from Theorems 4.4 and 5.7, together with $(E_n A E_n)^{-1} \to A^{-1}$ when $n \to \infty$.

With $\phi(t) = z - t$, $z \notin [-1, 1]$, we have already the convergence of the sequence of the diagonal Padé approximants to $f(z)$ at a fixed z. Uniform convergence on compacts could be deduced from uniform boundedness of $\|(E_n A E_n)^{-1}\|$ with respect to z and Vitali's theorem, but a more direct proof will follow.

In order to complete our information on the approximants, we consider first (7), i.e., the preceding problem with $A = M$, $\phi(t) = 1$, $M_n = E_n M E_n$. When n is large enough, $\| M_n^{-1} \| \leqslant C < \infty$, in particular, $\det M_n \neq 0$. From the end of Section 2, this settles (i).

For more details on the expansion (5) of P_n in terms of the Π_k's, let us consider the choice $h_n = 1$, and we add to (7) the equation $c(\Pi_n P_n) = \sum_{k=0}^n c(\Pi_n \Pi_k) \pi_{n-k,n} = 1/\pi_{0,n}$, so that we have

$$M_{n+1} x = y,$$

where $x = [\pi_{n,n}\pi_{0,n}, \pi_{n-1,n}\pi_{0,n}, \ldots, \pi_{0,n}\pi_{0,n}]^{\mathrm{T}}$ and $y = [0, 0, \ldots, 0, 1]^{\mathrm{T}}$. We know that the matrix M_{n+1} can be written $M_{n+1} = E_{n+1} T(g(\cos\theta)) E_{n+1} + E_{n+1} K E_{n+1}$, where K is compact. Turning the elements of M_{n+1} upside down and left to right, the equation becomes

$$\left(E_{n+1} T(g(\cos\theta)) E_{n+1} + J_{n+1}\right) z^{(n)} = u,$$

where $z^{(n)} = [\pi_{0,n}\pi_{0,n}, \ldots, \pi_{n,n}\pi_{0,n}]^{\mathrm{T}}$ and $u = [1, 0, 0, \ldots]^{\mathrm{T}}$. The Toeplitz part has been left unchanged, but now $J_{n+1} \to 0$ when $n \to \infty$! Indeed, the elements of $J_{n+1}x$ are the elements of $E_{n+1} K E_{n+1} [x_n, x_{n-1}, \ldots, x_0]^{\mathrm{T}}$, and this last factor converges weakly towards 0. Now, as the $z^{(n)}$'s are uniformly bounded, one finds easily that they converge weakly towards $(T(g(\cos\theta)))^{-1} u$, i.e.:

$$\pi_{j,n}\pi_{0,n} \xrightarrow[n \to \infty]{} (T^{-1}u)_j = \xi_j \eta_0,$$

or with a symmetric factorization $\rho_1 = \rho_2 = \rho$:

$$\pi_{j,n} \to \xi_j, \quad \text{when } n \to \infty, \quad j = 0, 1, \ldots$$

So, as far as high powers of z are concerned,

$$P_n(z) \sim \xi_0 \Pi_n(z) + \xi_1 \Pi_{n-1}(z) + \cdots.$$

As the recurrence relation coefficients for the P_n's depend only on the high powers of z, (17) follows then easily from (14), ($\xi_0 \neq 0$ geometric mean of $1/g$). From the uniform boundedness of the $z^{(n)}$'s and the non vanishing of ξ_0, limit of $\pi_{0,n}$ when $n \to \infty$, the uniform boundedness of $\sum_0^n |\pi_{i,n}|^2$ follows. Then, from (5) and $g \in L_\infty$, (15) is established.

Now, as $P_n/\Pi_n \sim \xi_0 + \xi_1 \Pi_{n-1}/\Pi_n + \xi_2 \Pi_{n-2}/\Pi_n \ldots$, and as the ratios Π_{n-i}/Π_n tend towards $(z - (z^2 - 1)^{1/2})^i$ when z is outside $[-1, 1]$ (from Poincaré's theorem [26], see also [9]), (iii) follows (the square root is such that $|z - (z^2 - 1)^{1/2}| < 1$).

Finally, (v) and (ii) follow from (6), Perron's theorem [26,9] and (15). □

Conclusion

Convergence of diagonal Padé approximants for a special class of functions has been established. The problem has been connected with convergence of projection methods of operators. Important results in Toeplitz operator theory have then been used. More flexible correspondences between Padé approximants and special operators should be welcome.

Acknowledgements

I wish to thank D.S. Lubinsky, P. Nevai and J. Nuttall for their informations and their interest.

References

[1] G.V. Ambarcumjan, On a reduction method for a class of Toeplitz matrices (Russian), *Mat. Issled.* **8** (1973) 161–169.

[2] G.A. Baker, Jr. and P. Graves-Morris, *Padé Approximants I* (Addison Wesley, Reading, MA, 1981).

[3] J. Bourgain, A problem of Douglas and Rudin on factorization, *Pac. J. Math.* **121** (1986) 47–50.

[4] C. Brezinski, *Padé-Type Approximation and General Orthogonal Polynomials* (Birkhäuser, Basel, 1980).

[5] M.G. de Bruin and H. van Rossum, Formal Padé approximation, *Nieuw Arch. Wiskunde* **23** (3) (1975) 115–130.

[6] R.G. Douglas, *Banach Algebra Techniques in Operator Theory* (Academic Press, New York, 1972).

[7] A. Draux, *Polynômes Orthogonaux Formels — Applications*, Lecture Notes Math. **974** (Springer, Berlin, 1983).

[8] E.N. Economou, *Green's Functions in Quantum Physics* (Springer, Berlin, 2nd ed., 1983).

[9] W. Gautschi, Computational aspects of three-term recurrence relations, *SIAM Rev.* **9** (1967) 24–82.

[10] W. Gautschi, Orthogonal polynomials-Constructive theory and applications, *J. Comp. Appl. Math.* **12&13** (1985) 61–76.

[11] P. Giannozzi, G. Grosso and G. Pastori Parravicini, The classical and generalized moment problem in the theory of relaxation, *Phys. Stat. Sol. (b)* **128** (1985) 643–652.

[12] J. Gilewicz, *Approximants de Padé*, Lecture Notes Math. **667** (Springer, Berlin 1978).

[13] U. Grenander and G. Szegö, *Toeplitz Forms and their Applications* (University of California Press, Berkeley, 1958).

[14] G. Grosso and G. Pastori Parravicini, Continued fractions in the theory of relaxation; Memory function methods in solid state physics, Chapter 3 and 4 (pp. 81–132 and 133–181) in: M.W. Evans, P. Grigolini, G. Pastori Parravicini, Eds., *Memory Function Approaches to Stochastic Problems in Condensed Matter*, Special Volume of *Adv. in Chem. Phys.* **62** (1985).

[15] L. Jacobsen and H. Waadeland, Modification of continued fractions, in: H. Werner and H.J. Bünger, Eds., *Padé Approximation and its Applications, Bad Honnef, 1983,* Lecture Notes Math. **1071** (Springer, Berlin, 1984) 176–196.

[16] L.V. Kantorovitch, Functional analysis and applied mathematics (transl. C.D. Bester), US Dept. Commerce, *NBS Rep.* **1509** (1952).

[17] L.V. Kantorovitch and G.P. Akilov, *Functional Analysis in Normed Spaces* (Pergamon Press, Oxford, 1964).

[18] J. Karlsson, Problem on continued fractions and complex measures, *Kong. Norske Vid. Selsk. Skr.* **1** (1983) 159.

[19] G. López Lagomasino, Survey on multipoint Padé approximation to Markov type meromorphic functions and asymptotic properties of the orthogonal polynomials generated by them, in: C. Brezinski et al., Eds., *Polynômes Orthogonaux et Applications; Laguerre Symposium, Bar-Le-Duc 1984*, Lecture Notes Math. **1171** (Springer, Berlin, 1985) 309–316.

[20] D.S. Lubinsky, Padé tables of a class of entire functions, *Proc. Amer. Math. Soc.* **94** (1985) 399–405.

[21] D.S. Lubinsky and E.B. Saff, Convergence of Padé approximants of partial theta functions and the Rogers–Szegö polynomials, *Constr. Approx.*, to appear.

[22] A. Magnus, Fractions continues généralisées et matrices infinies, *Bull. Soc. Math. Belg.* **29** *(ser. B)* (1977) 145–159.

[23] A. Máté, P. Nevai and V. Totik, Asymptotics for the ratio of leading coefficients of orthonormal polynomials on the unit circle, *Constr. Approx.* **1** (1985) 63–69.

[24] A. Mâté, P. Nevai and V. Totik, Extensions of Szegö's theory of orthogonal polynomials, II and III, *Constr. Approx.* **3** (1987) 51–72 and 73–96.
[25] J. Meinguet, On the Davis–Kahan–Weinberger solution of the norm-preserving dilation problem, *Numer. Math.* **49** (1986) 331–341.
[26] L.M. Milne-Thomson, *The Calculus of Finite Differences* (Macmillan, London, 1933).
[27] P. Nevai, Orthogonal polynomials, *Memoirs AMS* **18** (213) (1979).
[28] P. Nevai, Extensions of Szegö's theory of orthogonal polynomials, C. Brezinski et al., Eds., *Polynômes Orthogonaux et Applications; Laguerre Symposium, Bar-Le-Duc, 1984,* Lecture Notes Math. **1171** (Springer, Berlin, 1985) 230–238.
[29] P. Nevai, Private communication, 2/21/1968.
[30] J. Nuttall and S.R. Singh, Orthogonal polynomials and Padé approximants associated with a system of arcs, *J. Approx. Theory* **21** (1977) 1–42.
[31] J. Nuttall and C.J. Wherry, Gaussian integration for complex weight functions, *J. Inst. Math. Appl.* **21** (1978) 165–170.
[32] E.A. Rahmanov, On the asymptotics of the ratio of orthogonal polynomials, *Math. USSR Sb.* **32** (1977) 199–213; II *Math. USSR Sb.* **46** (1983) 105–117.
[33] H. Stahl, Orthogonal polynomials of complex-valued measures and the convergence of Padé approximants, in: G. Alexits and P. Turán, Eds., *Fourier Analysis and Approximation Theory, Vol. II* (North-Holland, Amsterdam, 1978) 771–788.
[34] H. Stahl, Divergence of diagonal Padé approximants and the asymptotic behavior of orthogonal polynomials associated with nonpositive measures, *Constr. Approx.* **1** (1985) 249–270.
[35] H. Stahl, On the divergence of certain Padé approximant and the behaviour of the associated orthogonal polynomials, in: C. Brezinski et al., Eds., *Polynômes Orthogonaux et Applications; Laguerre Symposium, Bar-Le-Duc, 1984*, Lecture Notes Math. **1771** (Springer, Berlin, 1985) 321–330.
[36] H. Stahl, Orthogonal polynomials with complex valued weight function I, II, *Constr. Approx.* **2** (1986) 241–251 and 253–261.
[37] H. Stahl, The convergence of Padé approximants to functions with branch point, preprint.
[38] J.W. Thron and H. Waadeland, Survey of modifications of continued fractions, in: W.B. Jones, W.J. Thron and H. Waadeland, Eds., *Analytic Theory of Continued Fractions*, Lecture Notes Math. **932** (Springer, Berlin, 1982) 38–66.
[39] H. Wallin, Interpolation by rational functions, in: H. Kurke et al., Eds., *Recent Trends in Mathematics, Reinhardsbrunn, 1982*, Teubner Texte zur Math. **50** (Teubner, Leipzig, 1982) 303–315.
[40] H. Widom, Toeplitz matrices, in: I.I. Hirschman, Jr., Ed., *Studies in Real and Complex Analysis* (Math. Assoc. Amer., 1965) 179–209.
[41] H. Widom, Toeplitz matrices and Toeplitz operators, 41 in: *Complex Analysis and its Applications, Vol. 1* (IAEA, Vienna, 1976) 319–341.
[42] I.C. Goldberg and I.A. Fel'dman, *Convolution Equations and Projection Methods for Their Solution*, Transl. Math. Monographs **41** (AMS, Providence, RI, 1974).