Padé and Hermite-Padé Approximation and Orthogonality

Walter Van Assche

1 September 2006

Abstract

We give a short introduction to Padé approximation (rational approximation to a function with close contact at one point) and to Hermite-Padé approximation (simultaneous rational approximation to several functions with close contact at one point) and show how orthogonality plays a crucial role. We give some insight into how logarithmic potential theory helps in describing the asymptotic behavior and the convergence properties of Padé and Hermite-Padé approximation.

MSC: 41A21, 41A28, 42C05, 33C45

Contents

Surveys in Approximation Theory
Volume 2, 2006. pp. 61–91.

ISSN 1555-578X

1 Padé approximation

1.1 Taylor polynomials

The general setup in approximation theory is that a function f is given and that one wants to approximate it with a *simpler* function g but in such a way that the difference between f and g is small. The advantage is that the simpler function g can be handled without too many difficulties but the disadvantage is that one loses some information since f and g are different.

In the setting of Padé approximation one starts with a function $f : \mathbb{C} \to \mathbb{C}$ for which a Taylor expansion is known in the neighborhood of a given point $a \in \mathbb{C}$, i.e.,

$$f(z) = \sum_{k=0}^{\infty} c_k (z-a)^k, \quad c_k = \frac{f^{(k)}(a)}{k!}. \tag{1.1}$$

The function f can not be computed exactly using this Taylor expansion since this requires an infinite number of additions (and multiplications). We can obtain a polynomial approximation by truncating after n terms. The corresponding approximations are Taylor polynomials given by

$$f_n(z) = \sum_{k=0}^{n-1} c_k (z-a)^k, \tag{1.2}$$

and these Taylor polynomials are therefore characterized by

$$f(z) - f_n(z) = \mathcal{O}((z-a)^n), \qquad z \to a. \tag{1.3}$$

This condition is a (confluent) interpolation condition which tells us that the difference $f - f_n$ has a zero of multiplicity n at the point a. We know an explicit formula for the Taylor polynomial, namely

$$f_n(z) = \sum_{k=0}^{n-1} \frac{f^{(k)}(a)}{k!} (z-a)^k,$$

and the error is given by

$$f(z) - f_n(z) = \sum_{k=n}^{\infty} \frac{f^{(k)}(a)}{k!} (z-a)^k.$$

If f is analytic in a domain Ω that contains a and if Γ is a closed contour in Ω encircling a once in the positive direction (counterclockwise), then Cauchy's formula gives

$$\frac{f^{(k)}(a)}{k!} = \frac{1}{2\pi i} \int_\Gamma \frac{f(\xi)}{(\xi - a)^{k+1}} \, d\xi,$$

and hence

$$f_n(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(\xi)}{\xi - a}\sum_{k=0}^{n-1}\left(\frac{z-a}{\xi-a}\right)^k \, d\xi = \frac{1}{2\pi i}\int_\Gamma \frac{f(\xi)}{\xi - z}\left[1 - \left(\frac{z-a}{\xi-a}\right)^n\right] \, d\xi.$$

The error then becomes

$$f(z) - f_n(z) = \frac{1}{2\pi i}\int_\Gamma \frac{f(\xi)}{\xi - z}\left(\frac{z-a}{\xi-a}\right)^n \, d\xi. \tag{1.4}$$

The convergence of f_n to f corresponds to the convergence of the Taylor series, and typically one has uniform convergence on closed disks $|z-a| \leq r$, where $r < \rho(f)$ and

$$\rho(f) := \sup\{R : f \text{ is analytic in } |z-a| < R\}$$

is the radius of convergence of the series in (1.1). Indeed, if we choose $\epsilon > 0$ such that $r + \epsilon < \rho(f)$ and if we take for Γ the circle $|\xi - a| = r + \epsilon$, then for $|z-a| \leq r$ we have from (1.4) by straightforward estimations

$$|f(z) - f_n(z)| \leq \max_{|\xi|=r+\epsilon} |f(\xi)| \left(\frac{r}{r+\epsilon}\right)^n \frac{1}{2\pi}\int_\Gamma \frac{|d\xi|}{|\xi - z|},$$

and since $r/(r+\epsilon) < 1$ we see that the right hand side converges to 0. So convergence is only guaranteed on disks with a radius less than the radius of convergence. The function f may be analytic in a larger domain (the radius of convergence depends on the singularity of f closest to a), but the Taylor approximation will not converge outside the disk with radius $\rho(f)$.

1.2 Padé approximants

Polynomials are not such a good class of functions if one wants to approximate functions with singularities because polynomials are entire functions without singularities. They are only useful up to the first singularity of f near a. Rational functions are the simplest functions with singularities. The idea is that the poles of the rational functions will move to the singularities of the function f, and hence the domain of convergence could be enlarged, and singularities of f may be discovered using the poles of the rational approximants.

The $[m, n]$ Padé approximant of f in a is the rational function Q_m/P_n, with Q_m a polynomial of degree $\leq m$ and P_n a polynomial of degree $\leq n$, for which we have the following interpolation condition at a:

$$f(z) - \frac{Q_m(z)}{P_n(z)} = \mathcal{O}((z-a)^{m+n+1}), \qquad z \to a. \tag{1.5}$$

The computation of the polynomials P_n and Q_m is not so easy from this interpolation condition since one first has the compute the Taylor expansion of Q_m/P_n and then equate the first $m+n+1$ Taylor coefficients to the first $m+n+1$ Taylor coefficients of f. Usually the Padé approximant is defined by linearizing the interpolation condition as

$$P_n(z)f(z) - Q_m(z) = \mathcal{O}((z-a)^{m+n+1}), \qquad z \to a. \tag{1.6}$$

For Padé approximation near infinity to a function of the form

$$f(z) = \sum_{k=0}^{\infty} \frac{c_k}{z^{k+1}},$$

one takes $m = n-1$ and the interpolation condition is

$$P_n(z)f(z) - Q_{n-1}(z) = \mathcal{O}(z^{-n-1}), \qquad z \to \infty,$$

(see Section 1.3). There is a degree of freedom since we can multiply both sides of (1.6) by a constant. Usually we normalize this by taking P_n **monic** (i.e., of the form $x^n + \cdots$) when this is possible, and this can only be done if P_n is of exact degree n. If we take P_n monic, then we can determine the n unknown coefficients a_k $(k = 1, \ldots, n)$ in

$$P_n(z) =: \sum_{k=0}^{n} a_k (z-a)^{n-k}, \qquad a_0 = 1, \tag{1.7}$$

by putting the coefficients of $(z-a)^k$ for $k = m+1, m+2, \ldots, m+n$ in the Taylor expansion of $P_n f$ equal to zero. The polynomial Q_m then corresponds to the Taylor polynomial of degree m of $P_n f$.

Here is another approach. Suppose f is analytic in a domain Ω that contains a. Again we take a contour Γ inside Ω encircling a once in the positive direction. Divide both sides of (1.6) by $(z-a)^{m+k+2}$ and integrate, to find

$$\frac{1}{2\pi i}\int_\Gamma \frac{P_n(z)f(z)}{(z-a)^{m+k+2}}\,dz - \frac{1}{2\pi i}\int_\Gamma \frac{Q_m(z)}{(z-a)^{m+k+2}}\,dz = \sum_{j=m+n+1}^{\infty} b_{n,j} \frac{1}{2\pi i}\int_\Gamma (z-a)^{j-m-k-2}\,dz,$$

where the $b_{n,j}$'s are the coefficients in the expansion of $P_n f - Q_m$ around a. The integral involving Q_m is zero for $k \geq 0$ since it is proportional to the $(m+k+1)$th derivative of Q_m, which is zero for $k \geq 0$. The sum on the right-hand side has a contribution only when $j = m+k+1$, but when $0 \leq k \leq n-1$ then $j \leq m+n$ and such indices do not appear in the sum. Hence the right hand side also vanishes for $k \leq n-1$. Therefore (1.6) implies that

$$\frac{1}{2\pi i}\int_\Gamma \frac{P_n(z)}{(z-a)^{m+k+2}} f(z)\,dz = 0, \qquad k = 0, 1, \ldots, n-1.$$

If we use the expansion (1.7) then this gives

$$\sum_{j=0}^{n} a_j \frac{1}{2\pi i}\int_\Gamma (z-a)^{n-j-m-k-2} f(z)\,dz = 0, \qquad k = 0, 1, 2, \ldots, n-1.$$

If we use the expansion (1.1) then

$$\frac{1}{2\pi i}\int_\Gamma (z-a)^{n-j-m-k-2} f(z)\,dz = c_{m-n+k+j+1},$$

so we get the system of equations

$$\begin{pmatrix} c_{m-n+1} & c_{m-n+2} & \cdots & c_{m+1} \\ c_{m-n+2} & c_{m-n+3} & \cdots & c_{m+2} \\ \vdots & \vdots & \cdots & \vdots \\ c_m & c_{m+1} & \cdots & c_{m+n} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}. \tag{1.8}$$

There is one degree of freedom here since we have $n+1$ unknowns and n (homogeneous) equations. The choice $a_0 = 1$ (if possible) gives the monic polynomial P_n, but sometimes another normalization will be used, as we will see later.

1.3 Orthogonality

From now on we will only consider Padé approximants near infinity. This can easily be obtained from Padé approximation near zero and the change of variable $z \mapsto 1/z$. Indeed, if g has a Taylor expansion

$$f^*(z) := \sum_{k=0}^{\infty} c_k z^k$$

near the origin, then $f(z) := g(1/z)/z$ has an expansion near infinity of the form

$$f(z) = \sum_{k=0}^{\infty} \frac{c_k}{z^{k+1}}. \tag{1.9}$$

Since $f(z) = \mathcal{O}(1/z)$, the only sensible choice of the degree in the rational approximation problem is to take $m = n-1$ so that $Q_m(z)/P_n(z)$ is also $\mathcal{O}(1/z)$. This situation occurs when f is of the form

$$f(z) = \int_{-\infty}^{\infty} \frac{d\mu(x)}{z-x},$$

i.e., when f is the Stieltjes transform (or Cauchy transform) of a positive measure μ on the real line. The Padé approximants near infinity can be obtained from the Padé approximants near zero in the following way. The $[n-1,n]$ Padé approximant Q_{n-1}^*/P_n^* for f^* near 0 has the interpolation condition

$$P_n^*(x) f^*(x) - Q_{n-1}^*(x) = \mathcal{O}(x^{2n}), \qquad x \to 0.$$

Change variables by setting $x = 1/z$ and divide both sides by z. Then

$$P_n^*(1/z) f(z) - \frac{1}{z} Q_{n-1}^*(1/z) = \mathcal{O}(z^{-2n-1}), \qquad z \to \infty.$$

In order to get polynomials, we multiply both sides by z^n. Then

$$P_n(z) f(z) - Q_{n-1}(z) = \mathcal{O}(z^{-n-1}), \qquad z \to \infty, \tag{1.10}$$

where $P_n(z) := z^n P_n^*(1/z)$ and $Q_{n-1}(z) := z^{n-1} Q_{n-1}^*(1/z)$ are obtained by reversing the polynomials P_n^* and Q_{n-1}^*. So the interpolation conditions at infinity are given by (1.10). The system of equations (1.8) for f^* and $m = n-1$ then changes to the system

$$\begin{pmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \vdots & \vdots & \cdots & \vdots \\ c_n & c_{n+1} & \cdots & c_{2n-1} \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \tag{1.11}$$

for the unknown coefficients of

$$P_n(z) := \sum_{k=0}^{n} a_k z^k.$$

Typically we will not be given the function f but rather the infinite sequence of coefficients $c_0, c_1, c_2, \ldots$ in the Laurent expansion of f. With this as input, we define a linear functional $\mathcal{L}$ on the linear space of polynomials by

$$\mathcal{L}(x^n) := c_n, \qquad n = 0, 1, 2, \ldots. \tag{1.12}$$

For a polynomial $p(x) = \sum_{k=0}^n a_k x^k$ we then have by linearity $\mathcal{L}(p) = \sum_{k=0}^n a_k c_k$. If we now look at the system of equations (1.11), then the coefficients of P_n satisfy the equations

$$\sum_{j=0}^n a_j c_{k+j} = 0, \qquad k = 0, 1, \ldots, n-1.$$

But this is equivalent to saying that

$$\mathcal{L}(x^k P_n(x)) = 0, \qquad k = 0, 1, \ldots, n-1. \tag{1.13}$$

Hence the polynomial P_n is orthogonal to all polynomials of degree less than n with respect to the linear functional $\mathcal{L}$. A very useful normalization of P_n is to require that in addition to (1.13) we also have

$$\mathcal{L}(P_n^2(x)) = 1.$$

This can always be done when the functional is positive. When the functional is not positive, then one imposes the extra condition $\mathcal{L}(P_n^2(x)) := h_n \neq 0$, so that $P_n/\sqrt{h_n}$ has norm one. Once the polynomial P_n is obtained, the remaining elements in the Padé approximation problem can be found explicitly in terms of P_n. Indeed, if we define

$$Q_{n-1}(z) := \mathcal{L}\left(\frac{P_n(z) - P_n(x)}{z - x}\right), \tag{1.14}$$

then, since $[P_n(z) - P_n(x)]/(z - x)$ is a polynomial of degree $n - 1$ in the variable z, Q_{n-1} is a polynomial of degree $n - 1$ and (1.14) is equivalent to

$$P_n(z)\mathcal{L}\left(\frac{1}{z-x}\right) - Q_{n-1}(z) = \mathcal{L}\left(\frac{P_n(x)}{z-x}\right).$$

The functional $\mathcal{L}$ was only defined on polynomials, but if we expand $1/(z - x)$ in a Laurent series, then (at least formally)

$$\mathcal{L}\left(\frac{1}{z-x}\right) = \mathcal{L}\left(\sum_{k=0}^\infty \frac{x^k}{z^{k+1}}\right) = \sum_{k=0}^\infty \frac{c_k}{z^{k+1}} = f(z),$$

so what needs to be shown is that

$$\mathcal{L}\left(\frac{P_n(x)}{z-x}\right) = \mathcal{O}(z^{-n-1}).$$

Using the Laurent series of $1/(z - x)$ we find

$$\mathcal{L}\left(\frac{P_n(x)}{z-x}\right) = \sum_{k=0}^\infty \frac{1}{z^{k+1}} \mathcal{L}(x^k P_n(x)),$$

and the orthogonality conditions (1.13) show that the terms with $k \leq n - 1$ vanish. The first term is therefore the term with $k = n$, which is $\mathcal{O}(1/z^{n+1})$. What we also learn from this proof is that the error in the Padé approximation problem is given explicitly by

$$P_n(z)f(z) - Q_{n-1}(z) = \mathcal{L}\left(\frac{P_n(x)}{z-x}\right), \tag{1.15}$$

which is again in terms of the polynomial P_n.

1.4 Moment problem

The linear functional $\mathcal{L}$ remains a bit mysterious. Obviously it is related to the function f, but we would like to know it somewhat more explicitly. The Riesz representation theorem tells us that every positive and bounded linear functional on the linear space of continuous functions with compact support on the real line can be represented by a finite positive measure μ on the real line as

$$\mathcal{L}(f) = \int_{-\infty}^{\infty} f(x)\, d\mu(x).$$

If we want to get convergence results for Padé approximation, then it would be convenient to work with a bounded and positive linear functional $\mathcal{L}$, which is represented by a finite positive measure μ. In that case

$$c_k = \int_{-\infty}^{\infty} x^k\, d\mu(x) \tag{1.16}$$

will be the moments of a positive measure μ and the function f is the Cauchy transform (Stieltjes transform) of the measure μ:

$$f(z) = \int_{-\infty}^{\infty} \frac{1}{z-x}\, d\mu(x).$$

Obviously not every infinite sequence $c_0, c_1, c_2, \ldots$ will lead to a positive and bounded linear functional. The moment problem is to obtain conditions on this infinite sequence $c_0, c_1, c_2, \ldots$ guaranteeing that they are the moments of a finite positive measure on the real line, as in (1.16). If the measure is supported on $(-\infty, \infty)$ then this is known as the **Hamburger moment problem**. If the measure is supported on the positive axis $[0, \infty)$ then we speak of the **Stieltjes moment problem**. If the measure is supported on a finite interval (usually $[0, 1]$), then this is known as the **Hausdorff moment problem**. A necessary and sufficient condition that the sequence $c_0, c_1, c_2, \ldots$ consist of moments of a positive measure on $(-\infty, \infty)$ is that all the **Hankel matrices**

$$\begin{pmatrix} c_0 & c_1 & \cdots & c_n \\ c_1 & c_2 & \cdots & c_{n+1} \\ \vdots & \vdots & \cdots & \vdots \\ c_n & c_{n+1} & \cdots & c_{2n} \end{pmatrix}$$

be positive definite. Observe that these are precisely the matrices appearing in (1.11).

From now on we will add one more restriction, namely that the measure be supported on a finite interval $[a, b]$. This simplifies our treatment by avoiding non-compactness of the support. So our function f will be a **Markov function**

$$f(z) = \int_a^b \frac{1}{z-x}\, d\mu(x),$$

and such a function is analytic in $\mathbb{C} \setminus [a, b]$. The singularities of this function therefore are located on the interval $[a, b]$. The linear functional in this case is given by

$$\mathcal{L}(g) = \int_a^b g(x)\, d\mu(x),$$

for every continuous function g on $[a, b]$. The denominator polynomials in the Padé approximation problem are orthogonal polynomials for the measure μ on the interval $[a, b]$, i.e.,

$$\int_a^b x^k P_n(x)\, d\mu(x) = 0, \qquad k = 0, 1, \ldots, n-1, \tag{1.17}$$

which we normalize so that they are orthonormal

$$\int_a^b P_n^2(x)\, d\mu(x) = 1. \tag{1.18}$$

The numerator polynomials are given by

$$Q_{n-1}(z) = \int_a^b \frac{P_n(z) - P_n(x)}{z - x}\, d\mu(x), \tag{1.19}$$

and the error is given by

$$P_n(z) f(z) - Q_{n-1}(z) = \int_a^b \frac{P_n(x)}{z - x}\, d\mu(x). \tag{1.20}$$

1.5 Zeros and poles

The idea of using rational approximation is that the singularities of the Padé approximant would give an idea of the singularities of the function f. This is indeed so when f is a Markov function. The singularities of the Padé approximant are poles at the zeros of P_n. A consequence of the orthogonality is that these zeros are simple and they all are on the open interval (a, b).

Theorem 1.1. *Suppose that the support of μ is an infinite set in $[a, b]$. Then all the zeros of P_n are simple and located on (a, b).*

Proof. Let $x_1, \ldots, x_m$ be the sign changes of P_n on (a, b), then obviously $m \leq n$ since each sign change is a zero. Suppose that $m < n$. Then introduce the polynomial $\pi_m(x) := (x - x_1)(x - x_2) \cdots (x - x_m)$. The function $P_n(x)\pi_m(x)$ does not change sign on $[a, b]$ and since the support of μ contains infinitely many points we conclude that

$$\int_a^b P_n(x)\pi_m(x)\, d\mu(x) \neq 0.$$

But P_n is orthogonal to all polynomials of degree $< n$, hence this integral is equal to 0. This contradiction implies that $m = n$. So P_n has n sign changes on (a, b), each a zero of P_n, hence each a simple zero of P_n, and P_n has no other zeros. $\square$

1.6 Convergence

When we study the convergence of the Padé approximants, we use (1.20) to find

$$f(z) - \frac{Q_{n-1}(z)}{P_n(z)} = \frac{1}{P_n(z)} \int_a^b \frac{P_n(x)}{z - x}\, d\mu(x).$$

Observe that

$$P_n(z) \int_a^b \frac{P_n(x)}{z-x}\, d\mu(x) = \int_a^b \frac{P_n(x)[P_n(z) - P_n(x)]}{z-x}\, d\mu(x) + \int_a^b \frac{P_n^2(x)}{z-x}\, d\mu(x).$$

The fraction $[P_n(z) - P_n(x)]/(z-x)$ is a polynomial of degree $n-1$ in the variable x, so by orthogonality the first integral on the right vanishes. This gives

$$P_n(z) \int_a^b \frac{P_n(x)}{z-x}\, d\mu(x) = \int_a^b \frac{P_n^2(x)}{z-x}\, d\mu(x),$$

and the error in Padé approximation becomes

$$f(z) - \frac{Q_{n-1}(z)}{P_n(z)} = \frac{1}{P_n^2(z)} \int_a^b \frac{P_n^2(x)}{z-x}\, d\mu(x). \tag{1.21}$$

This error contains two parts: on the one hand it contains the polynomial P_n for which we will describe the asymptotic behavior in the next subsection, and on the other hand it contains the integral

$$\int_a^b \frac{P_n^2(x)}{z-x}\, d\mu(x),$$

which is in fact a Markov function for the probability measure $P_n^2(x)\, d\mu(x)$ when P_n is the orthonormal polynomial. We can estimate this integral as follows. Suppose that z belongs to a compact set $K \subset \mathbb{C} \setminus [a,b]$. Then the distance

$$d_K := \inf\{|z-x| : z \in K, x \in [a,b]\}$$

between K and $[a,b]$ is strictly positive. Therefore we have

$$\left| \int_a^b \frac{P_n^2(x)}{z-x}\, d\mu(x) \right| \le \int_a^b \frac{P_n^2(x)}{|z-x|}\, d\mu(x) \le \frac{1}{d_K},$$

and this bound is independent of n. So the convergence of the Padé approximants is completely determined by the asymptotic behavior of P_n.

1.7 Asymptotic properties

In this subsection we describe the asymptotic behavior of $|P_n(z)|^{1/n}$ when $z \in K$, where K is a compact subset of $\mathbb{C} \setminus [a,b]$. If we denote the leading coefficient of P_n by $\gamma_n > 0$ and the zeros of P_n by $x_{1,n} < x_{2,n} < \cdots < x_{n,n}$, then

$$P_n(z) = \gamma_n \prod_{j=1}^n (z - x_{j,n}).$$

The asymptotic behavior thus requires knowing the behavior of γ_n and the asymptotic distribution of the zeros.

Let us first consider the asymptotic distribution of the zeros. Consider the discrete measure

$$\nu_n := \frac{1}{n} \sum_{j=1}^n \delta_{x_{j,n}},$$

where δ_c is the Dirac measure with mass 1 at the point c. The measure ν_n describes the distribution of the zeros of P_n. The asymptotic distribution corresponds to an investigation of the limit of this sequence of measures. All the zeros of P_n are on the interval $[a,b]$, so all the measures ν_n are probability measures on $[a,b]$. Helly's selection principle tells us that there will be a subsequence that converges weakly to a probability measure ν on $[a,b]$. This means that there is a subsequence (n_k) such that

$$\lim_{k\to\infty} \int_a^b g(x)\, d\nu_{n_k}(x) = \int_a^b g(x)\, d\nu(x),$$

for every continuous function g on $[a,b]$. For the monic polynomial $\hat{P}_n := P_n/\gamma_n$ we have

$$\frac{1}{n}\log|\hat{P}_n(z)| = \frac{1}{n}\sum_{j=1}^n \log|z-x_{j,n}| = \int_a^b \log|z-x|\, d\nu_n(x),$$

hence when $z \in K \subset \mathbb{C}\setminus[a,b]$, then the weak convergence implies that

$$\lim_{k\to\infty} |\hat{P}_{n_k}(z)|^{1/n_k} = \exp\left(\int_a^b \log|z-x|\, d\nu(x)\right).$$

Next, the leading coefficient γ_n solves a minimization problem:

Theorem 1.2. *We have*

$$\frac{1}{\gamma_n^2} = \min_{q_n(x)=x^n+\cdots} \int_a^b |q_n(x)|^2\, d\mu(x), \tag{1.22}$$

and the minimum is attained at the monic orthogonal polynomial $\hat{P}_n$.

Proof. We can write an arbitrary monic polynomial of degree n as $q_n = \hat{P}_n + \pi_{n-1}$, where π_{n-1} is a polynomial of degree $\le n-1$. We then have

$$\int_a^b |q_n(x)|^2\, d\mu(x) = \int_a^b |\hat{P}_n(x)|^2\, d\mu(x) + \int_a^b |\pi_{n-1}(x)|^2\, d\mu(x) + 2\int_a^b \hat{P}_n(x)\pi_{n-1}(x)\, d\mu(x).$$

The last integral vanishes because of orthogonality, so that

$$\min_{q_n(x)=x^n+\cdots} \int_a^b |q_n(x)|^2\, d\mu(x) = \int_a^b |\hat{P}_n(x)|^2\, d\mu(x) + \min_{\pi_{n-1}} \int_a^b |\pi_{n-1}(x)|^2\, d\mu(x).$$

The minimum on the right hand side is obtained by taking $\pi_{n-1} = 0$, so the minimum in (1.22) is obtained for the monic orthogonal polynomial. $\square$

Without going too much into details, this extremal problem for γ_n will in fact tell us that the asymptotic behavior of $\gamma_n^{1/n}$ and the asymptotic distribution of the zeros (the measure ν) are described by an equilibrium problem for (logarithmic) potentials. There is a unique probability measure μ_e on $[a,b]$ that minimizes the **logarithmic energy**

$$\int_a^b \int_a^b \log\frac{1}{|x-y|}\, d\sigma(x)d\sigma(y)$$

over all probability measures σ supported on $[a, b]$. This measure is given by

$$d\mu_e(x) = \frac{1}{\pi} \frac{dx}{\sqrt{(x-a)(b-x)}}, \quad x \in [a, b],$$

and has the property that its **logarithmic potential** satisfies

$$U(x; \mu_e) = \int_a^b \log \frac{1}{|x-y|} \, d\mu_e(y) = -\log \frac{b-a}{4}, \qquad x \in [a, b].$$

This **equilibrium measure** corresponds to the measure ν describing the asymptotic zero distribution when the orthogonality measure μ is sufficiently regular on $[a, b]$. A sufficient condition is that $\mu' > 0$ almost everywhere on $[a, b]$ (**Erdős-Turán condition**). Furthermore, we also have

$$\lim_{n\to\infty} \gamma_n^{1/n} = \frac{4}{b-a}.$$

Combining both results shows that when $\mu' > 0$ almost everywhere on $[a, b]$ we have

$$\lim_{n\to\infty} |P_n(z)|^{1/n} = \frac{4}{b-a} \exp\left(-\int_a^b \log \frac{1}{|z-x|} \, d\mu_e(x) \right).$$

When z is on the interval $[a, b]$, then the right hand side is equal to 1, but when z moves away from $[a, b]$, then the right hand side becomes > 1. On the equipotential curves

$$C_r := \{z \in \mathbb{C} \setminus [a, b] : \frac{4}{b-a} \exp\left(-\int_a^b \log \frac{1}{|z-x|} \, d\mu_e(x) \right) = r\}$$

with $r > 1$, we then conclude that

$$\lim_{n\to\infty} |f(z) - \frac{Q_{n-1}(z)}{P_n(z)}|^{1/n} = \frac{1}{r^2},$$

showing that we have exponential convergence.

2 Hermite-Padé approximation

Hermite-Padé approximation is simultaneous rational approximation to a vector of r functions $f_1, f_2, \ldots, f_r$, which are all given as Taylor series around a point $a \in \mathbb{C}$ and for which we require interpolation conditions at a. We will restrict our attention to Hermite-Padé approximation around infinity and impose interpolation conditions at infinity.

2.1 Definition

Suppose we are given r functions with Laurent expansions

$$f_j(z) = \sum_{k=0}^{\infty} \frac{c_{k,j}}{z^{k+1}}, \qquad j = 1, 2, \ldots, r.$$

There are basically two different types of Hermite-Padé approximation. First we will need multi-indices $\vec{n} = (n_1, n_2, \ldots, n_r) \in \mathbb{N}^r$ and their size $|\vec{n}| = n_1 + n_2 + \cdots + n_r$.

Definition 2.1 (Type I)**.** *Type I Hermite-Padé approximation to the vector* $(f_1, \ldots, f_r)$ *near infinity consists of finding a vector* $(A_{\vec{n},1}, \ldots, A_{\vec{n},r})$ *of polynomials and a polynomial* $B_{\vec{n}}$*, with* $A_{\vec{n},j}$ *of degree* $\leq n_j - 1$*, such that*

$$\sum_{j=1}^{r} A_{\vec{n},j}(z) f_j(z) - B_{\vec{n}}(z) = \mathcal{O}\left(\frac{1}{z^{|\vec{n}|}}\right), \qquad z \to \infty. \tag{2.1}$$

In type I Hermite-Padé approximation one wants to approximate a linear combination (with polynomial coefficients) of the r functions by a polynomial. This is often done for the vector of functions $f, f^2, \ldots, f^r$, where f is a given function. The solution of the equation

$$\sum_{j=1}^{r} A_{\vec{n},j}(z) \hat{f}^j(z) - B_{\vec{n}}(z) = 0$$

is an algebraic function which gives an algebraic approximant $\hat{f}$ for the function f.

Definition 2.2 (Type II)**.** *Type II Hermite-Padé approximation to the vector* $(f_1, \ldots, f_r)$ *near infinity consists of finding a polynomial* $P_{\vec{n}}$ *of degree* $\leq |\vec{n}|$ *and polynomials* $Q_{\vec{n},j}$ $(j = 1, 2, \ldots, r)$ *such that*

$$\begin{aligned} P_{\vec{n}}(z) f_1(z) - Q_{\vec{n},1}(z) &= \mathcal{O}\left(\frac{1}{z^{n_1+1}}\right), \qquad z \to \infty \\ &\vdots \\ P_{\vec{n}}(z) f_r(z) - Q_{\vec{n},r}(z) &= \mathcal{O}\left(\frac{1}{z^{n_r+1}}\right), \qquad z \to \infty. \end{aligned} \tag{2.2}$$

Type II Hermite-Padé approximation therefore corresponds to an approximation of each function f_j separately by rational functions *with a common denominator* $P_{\vec{n}}$. Combinations of type I and type II Hermite-Padé approximation are also possible.

2.2 Orthogonality

When we consider r Markov functions

$$f_j(z) = \int_{a_j}^{b_j} \frac{d\mu_j(x)}{z - x}, \qquad j = 1, 2, \ldots, r,$$

then Hermite-Padé approximation corresponds again to certain orthogonality conditions.

First, consider type I approximation. Multiply (2.1) by z^k and integrate over a contour Γ encircling all the intervals $[a_j, b_j]$ in the positive direction. Then

$$\frac{1}{2\pi i} \int_\Gamma \left(\sum_{j=1}^{r} z^k A_{\vec{n},j}(z) f_j(z) \right) dz - \frac{1}{2\pi i} \int_\Gamma z^k B_{\vec{n}}(z)\, dz = \sum_{\ell=|\vec{n}|}^{\infty} b_{\vec{n},\ell} \frac{1}{2\pi i} \int_\Gamma z^{k-\ell}\, dz,$$

where the $b_{\vec{n},\ell}$ are the coefficients of the Laurent expansion of the left hand side in (2.1). Cauchy's theorem implies

$$\frac{1}{2\pi i} \int_\Gamma z^k B_{\vec{n}}(z)\, dz = 0.$$

Furthermore, there is only a contribution on the right hand side when $\ell = k+1$, so when $k \leq |\vec{n}|-2$, then none of the terms in the infinite sum has a contribution. Therefore we see that

$$\frac{1}{2\pi i} \int_\Gamma \left(\sum_{j=1}^r z^k A_{\vec{n},j}(z) f_j(z) \right) dz = 0, \qquad 0 \leq k \leq |\vec{n}| - 2.$$

Now each f_j is a Markov function, so by changing the order of integration we get

$$\frac{1}{2\pi i} \int_\Gamma z^k A_{\vec{n},j}(z) f_j(z)\, dz = \int_{a_j}^{b_j} d\mu_j(x)\, \frac{1}{2\pi i} \int_\Gamma \frac{z^k A_{\vec{n},j}(z)}{z-x}\, dz.$$

Since Γ is a contour encircling $[a_j, b_j]$ we have that

$$\frac{1}{2\pi i} \int_\Gamma \frac{z^k A_{\vec{n},j}(z)}{z-x}\, dz = x^k A_{\vec{n},j}(x),$$

so that we get the following orthogonality conditions

$$\sum_{j=1}^r \int_{a_j}^{b_j} x^k A_{\vec{n},j}(x)\, d\mu_j(x) = 0, \qquad k = 0, 1, \ldots, |\vec{n}| - 2. \tag{2.3}$$

These are $|\vec{n}| - 1$ linear and homogeneous equations for the $|\vec{n}|$ coefficients of the r polynomials $A_{\vec{n},j}$ ($j = 1, 2, \ldots, r$), so that we can determine these polynomials up to a multiplicative factor, provided that the rank of the matrix in this system is $|\vec{n}| - 1$. If the solution is unique (up to a multiplicative factor), then we say that $\vec{n}$ is a **normal index for type I**. One can show that this is equivalent to the condition that the degree of each $A_{\vec{n},j}$ is exactly $n_j - 1$. Once the polynomial vector $(A_{\vec{n},1}, \ldots, A_{\vec{n},r})$ is determined, we can also find the remaining polynomial $B_{\vec{n}}$ which is given by

$$B_{\vec{n}}(z) = \sum_{j=1}^r \int_{a_j}^{b_j} \frac{A_{\vec{n},j}(z) - A_{\vec{n},j}(x)}{z-x}\, d\mu_j(x). \tag{2.4}$$

Indeed, with this definition of $B_{\vec{n}}$ we have

$$\sum_{j=1}^r A_{\vec{n},j}(z) f_j(z) - B_{\vec{n}}(z) = \sum_{j=1}^r \int_{a_j}^{b_j} \frac{A_{\vec{n},j}(x)}{z-x}\, d\mu_j(x). \tag{2.5}$$

If we use the expansion

$$\frac{1}{z-x} = \sum_{k=0}^\infty \frac{x^k}{z^{k+1}},$$

then the right hand side is

$$\sum_{k=0}^\infty \frac{1}{z^{k+1}} \sum_{j=1}^r \int_{a_j}^{b_j} x^k A_{\vec{n},j}(x)\, d\mu_j(x),$$

and the orthogonality conditions (2.3) show that the sum over k starts with $k = |\vec{n}| - 1$, hence the right hand side is $\mathcal{O}(z^{-|\vec{n}|})$, which is the order given in the definition of type I Hermite-Padé approximation.

Next, we consider type II approximation. Multiply (2.2) by z^k and integrate over a contour Γ encircling all the intervals $[a_j, b_j]$. Then

$$\frac{1}{2\pi i}\int_\Gamma z^k P_{\vec{n}}(z) f_j(z)\,dz - \frac{1}{2\pi i}\int_\Gamma z^k Q_{\vec{n},j}(z)\,dz = \sum_{\ell=n_j+1}^{\infty} b_{\vec{n},j,\ell}\frac{1}{2\pi i}\int_\Gamma z^{k-\ell}\,dz,$$

where the $b_{\vec{n},j,\ell}$ are the coefficients in the Laurent expansion of the left hand side of (2.2). Cauchy's theorem gives

$$\frac{1}{2\pi i}\int_\Gamma z^k Q_{\vec{n},j}(z)\,dz = 0,$$

and on the right hand side we only have a contribution when $\ell = k+1$. So for $k \leq n_j - 1$ none of the terms in the infinite sum contribute. Hence

$$\frac{1}{2\pi i}\int_\Gamma z^k P_{\vec{n}}(z) f_j(z)\,dz = 0, \qquad 0 \leq k \leq n_j - 1.$$

Interchanging the order of integration on the left hand side gives the orthogonality conditions

$$\begin{aligned} \int_{a_1}^{b_1} x^k P_{\vec{n}}(x)\,d\mu_1(x) &= 0, \qquad k = 0,1,\ldots,n_1-1, \\ &\vdots \\ \int_{a_r}^{b_r} x^k P_{\vec{n}}(x)\,d\mu_r(x) &= 0, \qquad k = 0,1,\ldots,n_r-1. \end{aligned} \tag{2.6}$$

This gives $|\vec{n}|$ linear and homogeneous equations for the $|\vec{n}|+1$ coefficients of $P_{\vec{n}}$, hence we can obtain the polynomial $P_{\vec{n}}$ up to a multiplicative factor, provided the matrix of coefficients has rank $|\vec{n}|$. In that case we call the index $\vec{n}$ **normal for type II**. One can show that this is equivalent to the condition that the degree of $P_{\vec{n}}$ be exactly $|\vec{n}|$. Once the polynomial $P_{\vec{n}}$ is determined, we can obtain the polynomials $Q_{\vec{n},j}$ by

$$Q_{\vec{n},j}(z) = \int_{a_j}^{b_j} \frac{P_{\vec{n}}(z) - P_{\vec{n}}(x)}{z-x}\,d\mu_j(x). \tag{2.7}$$

Indeed, with this expression for $Q_{\vec{n},j}$ we have

$$P_{\vec{n}}(z) f_j(z) - Q_{\vec{n},j}(z) = \int_{a_j}^{b_j} \frac{P_{\vec{n}}(x)}{z-x}\,d\mu_j(x), \tag{2.8}$$

and if we expand $1/(z-x)$, then the right hand side is of the form

$$\sum_{k=0}^{\infty} \frac{1}{z^{k+1}} \int_{a_j}^{b_j} x^k P_{\vec{n}}(x)\,d\mu_j(x),$$

and the orthogonality conditions (2.6) show that the infinite sum starts at $k = n_j$, which gives an expression of $\mathcal{O}(z^{-n_j-1})$, which is exactly what is required for type II Hermite-Padé approximation.

2.3 Angelesco systems

Angelesco [1] introduced an interesting system about which more can be said.

Definition 2.3. *An Angelesco system* $(f_1, f_2, \ldots, f_r)$ *consists of* r *Markov functions for which the intervals* (a_j, b_j) *are pairwise disjoint.*

All multi-indices are normal for type II in an Angelesco system. We will prove this by showing that the multiple orthogonal polynomial $P_{\vec{n}}$ has degree exactly equal to $|\vec{n}|$. In fact more is true, namely:

Theorem 2.1. *If* $(f_1, \ldots, f_r)$ *is an Angelesco system with measures* μ_j *that have infinitely many points in their support, then* $P_{\vec{n}}$ *has* n_j *simple zeros on* (a_j, b_j) *for* $j = 1, \ldots, r$.

Proof. Let $x_1, \ldots, x_m$ be the sign changes of $P_{\vec{n}}$ on (a_j, b_j). Suppose that $m < n_j$ and let $\pi_m(x) := (x - x_1)\cdots(x - x_m)$. Then $P_{\vec{n}}\pi_m$ does not change sign on $[a_j, b_j]$. Since the support of μ_j has infinitely many points, we have

$$\int_{a_j}^{b_j} P_{\vec{n}}(x)\pi_m(x)\, d\mu_j(x) \neq 0.$$

However, the orthogonality (2.6) implies that $P_{\vec{n}}$ is orthogonal to all polynomials of degree $\leq n_j - 1$ with respect to the measure μ_j on $[a_j, b_j]$, so that the integral is zero. This contradiction implies that $m \geq n_j$, and hence $P_{\vec{n}}$ has at least n_j zeros on (a_j, b_j). This holds for every j, and since the intervals (a_j, b_j) are disjoint this gives at least $|\vec{n}|$ zeros on the real line. But the degree of $P_{\vec{n}}$ is $\leq |\vec{n}|$, hence $P_{\vec{n}}$ has exactly n_j simple zeros on (a_j, b_j). □

The polynomial $P_{\vec{n}}$ can therefore be factored as

$$P_{\vec{n}}(x) = q_{n_1}(x)q_{n_2}(x)\cdots q_{n_r}(x),$$

where each q_{n_j} is a polynomial of degree n_j with its zeros on (a_j, b_j). The orthogonality (2.6) then gives

$$\int_{a_j}^{b_j} x^k q_{n_j}(x) \prod_{i\neq j} q_{n_i}(x)\, d\mu_j(x) = 0, \qquad k = 0, 1, \ldots, n_j - 1. \tag{2.9}$$

The product $\prod_{i\neq j} q_{n_i}(x)$ does not change sign on (a_j, b_j), hence (2.9) shows that q_{n_j} is an ordinary orthogonal polynomial of degree n_j on the interval $[a_j, b_j]$ with respect to the measure $\prod_{i\neq j} |q_{n_i}(x)|\, d\mu_j(x)$. The measure depends on the multi-index $\vec{n}$.

2.4 Algebraic Chebyshev systems

A Chebyshev system $(\varphi_1, \ldots, \varphi_n)$ on $[a, b]$ is a linearly independent system of n functions such that every nontrivial linear combination $\sum_{k=1}^n a_k\varphi_k$ has at most $n - 1$ zeros on $[a, b]$. This is equivalent to the condition that

$$\det \begin{pmatrix} \varphi_1(x_1) & \varphi_1(x_2) & \cdots & \varphi_1(x_n) \\ \varphi_2(x_1) & \varphi_2(x_2) & \cdots & \varphi_2(x_n) \\ \vdots & \vdots & \cdots & \vdots \\ \varphi_n(x_1) & \varphi_n(x_2) & \cdots & \varphi_n(x_n) \end{pmatrix} \neq 0$$

for every choice of n distinct points $x_1, \ldots, x_n \in [a,b]$. Indeed, when $x_1, \ldots, x_n$ are such that the determinant is zero, then there is a linear combination of the rows that gives a zero row, but this means that for this linear combination $\sum_{k=1}^n a_k \varphi_k$ has zeros at $x_1, \ldots, x_n$, giving n zeros, which is not allowed.

Definition 2.4. *A system $(f_1, \ldots, f_r)$ is an algebraic Chebyshev system (AT system) for the index $\vec{n}$ if each f_j is a Markov function on the same interval $[a,b]$ with a measure $w_j(x)\,d\mu(x)$, where μ has infinite support and the w_j are such that*

$$\{w_1, xw_1, \ldots, x^{n_1-1}w_1, w_2, xw_2, \ldots, x^{n_2-1}w_2, \ldots, w_r, xw_r, \ldots, x^{n_r-1}w_r\} \tag{2.10}$$

is a Chebyshev system on $[a,b]$.

Theorem 2.2. *Suppose $\vec{n}$ is a multi-index such that $(f_1, \ldots, f_r)$ is an AT system on $[a,b]$ for every index $\vec{m}$ for which $m_j \leq n_j$ $(1 \leq j \leq r)$. Then $P_{\vec{n}}$ has $|\vec{n}|$ zeros on (a,b) and hence $\vec{n}$ is a normal index for type II.*

Proof. Let $x_1, \ldots, x_m$ be the sign changes of $P_{\vec{n}}$ on (a,b) and suppose that $m < |\vec{n}|$. We can then find a multi-index $\vec{m}$ such that $|\vec{m}| = m$ and $m_j \leq n_j$ for every $1 \leq j \leq r$ and $m_k < n_k$ for some $1 \leq k \leq r$. Consider the interpolation problem where we want to find a function

$$L(x) = \sum_{j=1}^r q_j(x) w_j(x),$$

where q_j is a polynomial of degree $m_j - 1$ if $j \neq k$ and q_k a polynomial of degree m_k that satisfies

$$\begin{aligned} L(x_j) &= 0, \qquad j = 1, ..., m, \\ L(x_0) &= 1, \qquad \text{for some other point } x_0 \in [a,b]. \end{aligned}$$

The function L is a linear combination of

$$\{w_1, xw_1, \ldots, x^{m_1-1}w_1, \ldots, w_k, xw_k, \ldots, x^{n_k}w_k, \ldots w_r, xw_r, \ldots, x^{m_r-1}w_r\}$$

and this is, by assumption, a Chebyshev system. This interpolation problem has a unique solution since it involves a Chebyshev system of basis functions. The function L has, by construction, m zeros and the Chebyshev system has $m+1$ basis functions, so L can have at most m zeros on $[a,b]$ and each zero is a sign change (see, e.g., [23, pp. 20–21]). Hence $P_{\vec{n}}L$ does not change sign on $[a,b]$. Since μ has infinite support, we thus have

$$\int_a^b L(x) P_{\vec{n}}(x)\,d\mu(x) \neq 0.$$

But the orthogonality (2.6) gives

$$\int_a^b q_j(x) P_{\vec{n}}(x) w_j(x)\,d\mu(x) = 0, \qquad j = 1, 2, \ldots, r,$$

and this contradiction implies that $P_{\vec{n}}$ has $|\vec{n}|$ simple zeros on (a,b). □

We have a similar result for type I Hermite-Padé approximation:

Theorem 2.3. *Suppose $\vec{n}$ is a multi-index such that $(f_1, \ldots, f_r)$ is an AT system on $[a, b]$ for every index $\vec{m}$ for which $m_j \leq n_j$ $(1 \leq j \leq r)$. Then $\sum_{j=1}^r A_{\vec{n},j} w_j$ has $|\vec{n}| - 1$ zeros on (a, b) and $\vec{n}$ is a normal index for type I.*

Proof. Let $x_1, \ldots, x_m$ be the sign changes of $\sum_{j=1}^r A_{\vec{n},j} w_j$ on (a, b) and suppose that $m < |\vec{n}| - 1$. Let π_m be the monic polynomial with these points as zeros. Then $\pi_m \sum_{j=1}^r A_{\vec{n},j} w_j$ does not change sign on $[a, b]$ and hence

$$\int_a^b \pi_m(x) \sum_{j=1}^r A_{\vec{n},j}(x) w_j(x)\, d\mu(x) \neq 0.$$

But the orthogonality conditions (2.3) indicate that this integral is zero. This contradiction implies that $m \geq |\vec{n}| - 1$. The sum $\sum_{j=1}^r A_{\vec{n},j} w_j$ is a linear combination of the Chebyshev system (2.10), hence it has at most $|\vec{n}| - 1$ zeros on $[a, b]$. Therefore we see that $m = |\vec{n}| - 1$. To see that the index $\vec{n}$ is normal for type I, we assume that for some k with $1 \leq k \leq r$ the degree of $A_{\vec{n},k}$ is less than $n_k - 1$. Then $\sum_{j=1}^r A_{\vec{n},j} w_j$ is a linear combination of the Chebyshev system (2.10) from which the function $x^{n_k - 1} w_k$ is removed. This is still a Chebyshev system by assumption, and hence this linear combination has at most $|\vec{n}| - 2$ zeros on $[a, b]$. But this contradicts our previous observation that it has $|\vec{n}| - 1$ zeros. Therefore every $A_{\vec{n},j}$ has degree exactly $n_j - 1$, so that the index $\vec{n}$ is normal. □

2.5 Nikishin systems

A special construction, suggested by Nikishin [26], gives an AT system that can be handled in some detail. The construction is by induction. A **Nikishin system of order 1** is a Markov function $f_{1,1}$ for a measure μ_1 on the interval $[a_1, b_1]$. A **Nikishin system of order 2** is a vector of Markov functions $(f_{1,2}, f_{2,2})$ on $[a_2, b_2]$ such that

$$f_{1,2}(z) = \int_{a_2}^{b_2} \frac{d\mu_2(x)}{z - x}, \quad f_{2,2}(z) = \int_{a_2}^{b_2} f_{1,1}(x) \frac{d\mu_2(x)}{z - x},$$

where $f_{1,1}$ is a Nikishin system of order 1 on $[a_1, b_1]$ and $(a_1, b_1) \cap (a_2, b_2) = \emptyset$. In general we have

Definition 2.5. *A* **Nikishin system of order** *r consists of r Markov functions $(f_{1,r}, \ldots, f_{r,r})$ on $[a_r, b_r]$ such that*

$$f_{1,r}(z) = \int_{a_r}^{b_r} \frac{d\mu_r(x)}{z - x}, \tag{2.11}$$

$$f_{j,r}(z) = \int_{a_r}^{b_r} f_{j-1,r-1}(x) \frac{d\mu_r(x)}{z - x}, \qquad j = 2, \ldots, r, \tag{2.12}$$

where $(f_{1,r-1}, \ldots, f_{r-1,r-1})$ is a Nikishin system of order $r - 1$ on $[a_{r-1}, b_{r-1}]$ and (a_r, b_r) and (a_{r-1}, b_{r-1}) are disjoint.

For a Nikishin system of order r one knows that the multi-indices $\vec{n}$ with $n_1 \geq n_2 \geq \cdots \geq n_r$ are normal (the system is an AT-system for these indices), but it is an open problem whether every multi-index is normal (for $r > 2$; for $r = 2$ it has been proved that every multi-index is normal).

What can be said about type II Hermite-Padé approximation for $r = 2$? Recall (2.8) for the function $f_{1,2}$:

$$P_{n_1,n_2}(y)f_{1,2}(y) - Q_{n_1,n_2;1}(y) = \int_{a_2}^{b_2} \frac{P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x).$$

Multiply both sides by y^k, with $k \leq n_1$. Then the right hand side is

$$\int_{a_2}^{b_2} \frac{y^k P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x) = \int_{a_2}^{b_2} \frac{(y^k - x^k)P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x) + \int_{a_2}^{b_2} \frac{x^k P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x).$$

Clearly $(y^k - x^k)/(y-x)$ is a polynomial in x of degree $k-1 \leq n_1 - 1$ hence the first integral on the right vanishes because of the orthogonality (2.6). Integrate over the variable $y \in [a_1, b_1]$ with respect to the measure μ_1. Then we find for $k \leq n_1$

$$\int_{a_1}^{b_1} [P_{n_1,n_2}(y)f_{1,2}(y) - Q_{n_1,n_2;1}(y)]y^k\, d\mu_1(y) = \int_{a_1}^{b_1}\int_{a_2}^{b_2} \frac{x^k P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x)\, d\mu_1(y).$$

Change the order of integration on the right hand side. Then

$$\int_{a_1}^{b_1} [P_{n_1,n_2}(y)f_{1,2}(y) - Q_{n_1,n_2;1}(y)]y^k\, d\mu_1(y) = -\int_{a_2}^{b_2} x^k P_{n_1,n_2}(x) f_{1,1}(x)\, d\mu_2(x)$$

and this is zero for $k \leq n_2 - 1$. Hence if $n_2 \leq n_1 + 1$ then the expression $P_{n_1,n_2}f_{1,2} - Q_{n_1,n_2;1}$ is orthogonal to all polynomials of degree $\leq n_2 - 1$ on $[a_1, b_1]$. This implies that $P_{n_1,n_2}f_{1,2} - Q_{n_1,n_2;1}$ has at least n_2 zeros on (a_1, b_1) using an argument similar to what we have been using earlier. Let R_{n_2} be the monic polynomial with n_2 of these zeros on (a_1, b_1). Then $[P_{n_1,n_2}f_{1,2} - Q_{n_1,n_2;1}]/R_{n_2}$ is an analytic function on $\mathbb{C} \setminus [a_2, b_2]$, which has the representation

$$\frac{P_{n_1,n_2}(y)f_{1,2}(y) - Q_{n_1,n_2;1}(y)}{R_{n_2}(y)} = \frac{1}{R_{n_2}(y)} \int_{a_2}^{b_2} \frac{P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x).$$

Multiply both sides by y^k and integrate over a contour Γ encircling the interval $[a_2, b_2]$ in the positive direction, but with all the zeros of R_{n_2} outside Γ. Then

$$\frac{1}{2\pi i}\int_\Gamma y^k \frac{P_{n_1,n_2}(y)f_{1,2}(y) - Q_{n_1,n_2;1}(y)}{R_{n_2}(y)}\, dy = \frac{1}{2\pi i}\int_\Gamma \frac{y^k}{R_{n_2}(y)} \frac{P_{n_1,n_2}(x)}{y-x}\, d\mu_2(x)\, dy.$$

If we interchange the order of integration on the right hand side and use Cauchy's theorem, then this gives the integral

$$\int_{a_2}^{b_2} x^k P_{n_1,n_2}(x)\, \frac{d\mu_2(x)}{R_{n_2}(x)}.$$

According to the interpolation condition (2.2), the integrand on the left hand side is of the order $\mathcal{O}(y^{k-n_1-n_2-1})$, so if we use Cauchy's theorem for the exterior of Γ, then we see that the integral vanishes for $k \leq n_1 + n_2 - 1$. Hence we get

$$\int_{a_2}^{b_2} x^k P_{n_1,n_2}(x)\, \frac{d\mu_2(x)}{R_{n_2}(x)} = 0, \qquad k = 0, 1, \ldots, n_1 + n_2 - 1. \tag{2.13}$$

This shows that P_{n_1,n_2} is an ordinary orthogonal polynomial on $[a_2, b_2]$ with respect to the measure $d\mu_2/R_{n_2}$. Observe that $(a_1, b_1) \cap (a_2, b_2) = \emptyset$ implies that R_{n_2} does not change sign on $[a_2, b_2]$. Finally, we have

$$\begin{aligned}\int_{a_2}^{b_2} \frac{P_{n_1,n_2}^2(x)}{y-x} \frac{d\mu_2(x)}{R_{n_2}(x)} &= \int_{a_2}^{b_2} P_{n_1,n_2}(x) \frac{P_{n_1,n_2}(x) - P_{n_1,n_2}(y)}{y-x} \frac{d\mu_2(x)}{R_{n_2}(x)} \\ &+ P_{n_1,n_2}(y) \int_{a_2}^{b_2} \frac{P_{n_1,n_2}(x)}{y-x} \frac{d\mu_2(x)}{R_{n_2}(x)} \\ &= P_{n_1,n_2}(y) \int_{a_2}^{b_2} \frac{P_{n_1,n_2}(x)}{y-x} \frac{d\mu_2(x)}{R_{n_2}(x)},\end{aligned}$$

since $[P_{n_1,n_2}(y) - P_{n_1,n_2}(x)]/(y-x)$ is a polynomial in x of degree $n_1 + n_2 - 1$ and because of the orthogonality (2.13). Hence

$$P_{n_1,n_2}(y) f_{1,2}(y) - Q_{n_1,n_2;1}(y) = \frac{R_{n_2}(y)}{P_{n_1,n_2}(y)} \int_{a_2}^{b_2} \frac{P_{n_1,n_2}^2(x)}{y-x} \frac{d\mu_2(x)}{R_{n_2}(x)}. \tag{2.14}$$

Both sides of the equation have zeros at the zeros of R_{n_2}, but there will not be any other zeros on $[a_1, b_1]$ since the integral on the right hand side has constant sign.

2.6 Asymptotic properties and convergence

We restrict ourselves to the case $r = 2$, but the general case $r > 1$ can be treated in a similar way (with a bit more work). The asymptotic properties of the multiple orthogonal polynomials and the convergence of the Hermite-Padé approximants are handled by trying to put everything into terms of ordinary orthogonal polynomials.

2.6.1 Angelesco systems

The type II multiple orthogonal polynomial can be factored as $P_{n_1,n_2} = q_{n-1} q_{n-2}$, where q_{n_1} has n_1 zeros on (a_1, b_1) and q_{n_2} has n_2 zeros on (a_2, b_2). From (2.8) we get

$$f_1(z) - \frac{Q_{n_1,n_2;1}(z)}{P_{n_1,n_2}(z)} = \frac{1}{q_{n_1}(z) q_{n_2}(z)} \int_{a_1}^{b_1} \frac{q_{n_1}(x)}{z-x} q_{n_2}(x)\, d\mu_1(x).$$

We saw that q_{n_1} is an orthogonal polynomial of degree n_1 on $[a_1, b_1]$ for the measure $|q_{n_2}(x)|\, d\mu_1(x)$, so we can write

$$\int_{a_1}^{a_2} \frac{q_{n_1}(x)}{z-x} q_{n_2}(x)\, d\mu_1(x) = \frac{1}{q_{n_1}(z)} \int_{a_1}^{b_1} \frac{q_{n_1}^2(x)}{z-x} q_{n_2}(x)\, d\mu_1(x)$$

as we did earlier in Section 1.6. This gives

$$f_1(z) - \frac{Q_{n_1,n_2;1}(z)}{P_{n_1,n_2}(z)} = \frac{1}{q_{n_1}^2(z) q_{n_2}(z)} \int_{a_1}^{a_2} \frac{q_{n_1}^2(x)}{z-x} q_{n_2}(x)\, d\mu_1(x).$$

From here we get the estimate

$$\left| f_1(z) - \frac{Q_{n_1,n_2;1}(z)}{P_{n_1,n_2}(z)} \right| \le \frac{1}{|q_{n_1}(z)|^2 |q_{n_2}(z)|} \frac{1}{d_1} \int_{a_1}^{b_1} q_{n_1}^2(x)\, |q_{n_2}(x)|\, d\mu_1(x),$$

where d_1 is the distance between z and $[a_1, b_1]$. If P_{n_1,n_2} is normalized so that it is monic, then we can take both q_{n_1} and q_{n_2} monic and we get

$$\left| f_1(z) - \frac{Q_{n_1,n_2;1}(z)}{P_{n_1,n_2}(z)} \right| \leq \frac{1}{d_1 \gamma_{n_1;1}^2 |q_{n_1}(z)|^2 |q_{n_2}(z)|},$$

where

$$\begin{aligned} \frac{1}{\gamma_{n_1;1}^2} &= \int_{a_1}^{b_1} q_{n_1}^2(x)\, |q_{n_2}(x)|\, d\mu_1(x) \\ &= \min_{\pi_{n_1}(x)=x^{n_1}+\cdots} \int_{a_1}^{b_1} \pi_{n_1}^2(x)\, |q_{n_2}(x)|\, d\mu_1(x). \end{aligned} \tag{2.15}$$

A similar reasoning holds for the rational approximation to f_2 and gives

$$\left| f_2(z) - \frac{Q_{n_1,n_2;2}(z)}{P_{n_1,n_2}(z)} \right| \leq \frac{1}{d_2 \gamma_{n_2;2}^2 |q_{n_2}(z)|^2 |q_{n_1}(z)|},$$

where d_2 is the distance of z to $[a_2, b_2]$ and

$$\begin{aligned} \frac{1}{\gamma_{n_1;2}^2} &= \int_{a_2}^{b_2} q_{n_2}^2(x)\, |q_{n_1}(x)|\, d\mu_2(x) \\ &= \min_{\pi_{n_2}(x)=x^{n_2}+\cdots} \int_{a_2}^{b_2} \pi_{n_2}^2(x)\, |q_{n_1}(x)|\, d\mu_2(x). \end{aligned} \tag{2.16}$$

The convergence of these rational approximants is therefore given in terms of the asymptotic behavior of $|q_{n_1}(z)|$, $|q_{n_2}(z)|$ and the constants $\gamma_{n_1;1}$ and $\gamma_{n_2;2}$. These polynomials (and their zeros) interact with each other: the polynomial q_{n_1} is an orthogonal polynomial for a measure that contains q_{n_2} as a factor, and q_{n_2} is an orthogonal polynomial for a measure that contains q_{n_1} as a factor. Let

$$\nu_{n_1;1} := \frac{1}{n_1} \sum_{j=1}^{n_1} \delta_{x_{j,n_1}}, \quad \nu_{n_2;2} := \frac{1}{n_2} \sum_{j=1}^{n_2} \delta_{y_{j,n_2}},$$

where x_{j,n_1} are the zeros of q_{n_1} and y_{j,n_2} are the zeros of q_{n_2}. Then $(\nu_{n_1;1})$ is a sequence of probability measures on $[a_1, b_1]$ and $(\nu_{n_2;2})$ is a sequence of probability measures on $[a_2, b_2]$. Helly's selection principle guarantees that there are weakly converging subsequences with limits ν_1 on $[a_1, b_1]$ and ν_2 on $[a_2, b_2]$. The minimization problems (2.15) and (2.16) lead to an extremal problem in potential theory for two probability measures. The integral in (2.15) is approximately of the form

$$\int_{a_1}^{b_1} \exp\left[-2n_1 U(x;\nu_1) - n_2 U(x;\nu_2)\right]\, d\mu_1(x)$$

where $U(x;\nu)$ is the logarithmic potential of ν

$$U(x;\nu) = \int \log \frac{1}{|x-y|}\, d\nu(y),$$

and the integral in (2.16) is approximately of the form

$$\int_{a_2}^{b_2} \exp\left[-2n_2 U(x;\nu_2) - n_1 U(x;\nu_1)\right]\, d\mu_2(x).$$

We want to minimize both integrals over all pairs of probability measures (ν_1, ν_2), where the first measure is supported on $[a_1, b_1]$ and the second measure on $[a_2, b_2]$. If $n_1/(n_1 + n_2) \to p$ and $n_2/(n_1 + n_2) \to q$ (so that $p + q = 1$), and if the measures μ_1 and μ_2 are sufficiently regular (e.g., $\mu_1' > 0$ almost everywhere on $[a_1, b_1]$ and $\mu_2' > 0$ almost everywhere on $[a_2, b_2]$) then the solution of the extremal problem satisfies

$$\begin{aligned} 2pU(x;\nu_1) + qU(x;\nu_2) &= \ell_1, \qquad x \in \mathrm{supp}(\nu_1) \subset [a_1, b_1], && (2.17)\\ pU(x;\nu_1) + 2qU(x;\nu_2) &= \ell_2, \qquad x \in \mathrm{supp}(\nu_2) \subset [a_2, b_2], && (2.18)\end{aligned}$$

where the ℓ_j are constants that act as Lagrange multipliers. For this extremal problem it is possible that the support of ν_1 is not the full interval $[a_1, b_1]$ and the support of ν_2 can be a subset of $[a_2, b_2]$. This is a consequence of the interaction: the zeros of q_{n_1} are repelling the zeros of q_{n_2} and vice versa. The variational conditions (2.17)–(2.18) have to be supplemented with

$$\begin{aligned} 2pU(x;\nu_1) + qU(x;\nu_2) &\geq \ell_1, \qquad x \in [a_1, b_1] \setminus \mathrm{supp}(\nu_1),\\ pU(x;\nu_1) + 2qU(x;\nu_2) &\geq \ell_2, \qquad x \in [a_2, b_2] \setminus \mathrm{supp}(\nu_2).\end{aligned}$$

The Lagrange multipliers ℓ_1, ℓ_2 appear in the asymptotics of $\gamma_{n_1;1}$ and $\gamma_{n_2;2}$ as

$$\lim_{n_1+n_2\to\infty} \gamma_{n_1;1}^{2/(n_1+n_2)} = \exp(\ell_1), \qquad \lim_{n_1+n_2\to\infty} \gamma_{n_2;2}^{2/(n_1+n_2)} = \exp(\ell_2).$$

Our conclusion is that the convergence to the first function f_1 is determined by level curves

$$C_r := \{z : \exp[2pU(z;\nu_1) + qU(z;\nu_2) - \ell_1] = r\}$$

with $r < 1$, on which we have

$$\lim_{n_1+n_2\to\infty} \left| f_1(z) - \frac{Q_{n_1,n_2;1}(z)}{P_{n_1,n_2}(z)} \right|^{1/(n_1+n_2)} = r,$$

and the convergence to the second function f_2 is determined by level curves

$$D_r := \{z : \exp[pU(z;\nu_1) + 2qU(z;\nu_2) - \ell_2] = r\}$$

with $r < 1$, on which we have

$$\lim_{n_1+n_2\to\infty} \left| f_2(z) - \frac{Q_{n_1,n_2;2}(z)}{P_{n_1,n_2}(z)} \right|^{1/(n_1+n_2)} = r.$$

Observe that $\mathrm{supp}(\nu_1) \subset C_1$ and $\mathrm{supp}(\nu_2) \subset D_1$, so we don't expect exponential convergence on these sets. On the remaining part of $[a_1, b_1]$ (and $[a_2, b_2]$) we get values $r \geq 1$, so we get even worse behavior there. This is caused by the fact that on these parts of the intervals there will not be enough zeros of the multiple orthogonal polynomial to simulate the singularities of the functions f_1 and f_2.

2.6.2 Nikishin systems

The analysis for Nikishin systems is similar but leads to a different extremal problem for potentials. We now start from (2.14) which gives

$$\left| f_{1,2}(y) - \frac{Q_{n_1,n_2;1}(y)}{P_{n_1,n_2}(y)} \right| \leq \frac{|R_{n_2}(y)|}{|P_{n_1,n_2}(y)|^2} \frac{1}{d_2} \int_{a_2}^{b_2} P_{n_1,n_2}^2(x) \frac{d\mu_2(x)}{|R_{n_2}(x)|}, \tag{2.19}$$

where d_2 is the distance from y to $[a_2, b_2]$. Now we have that P_{n_1,n_2} is a (monic) orthogonal polynomial on $[a_2, b_2]$ for the measure $d\mu_2/|R_{n_2}|$, so we have

$$\begin{aligned} \frac{1}{\gamma_{n_1,n_2}^2} &= \int_{a_2}^{b_2} P_{n_1,n_2}^2(x) \frac{d\mu_2(x)}{|R_{n_2}(x)|} \\ &= \min_{\pi_{n_1+n_2}(x)=x^{n_1+n_2}+\cdots} \int_{a_2}^{b_2} \pi_{n_1,n_2}^2(x) \frac{d\mu_2(x)}{|R_{n_2}(x)|}. \end{aligned} \tag{2.20}$$

The polynomial R_{n_2} has its zeros on $[a_1, b_1]$ and in fact is a monic orthogonal polynomial on $[a_1, b_1]$ for the measure

$$\frac{P_{n_1,n_2} f_{1,2} - Q_{n_1,n_2;1}}{R_{n_2}} \, d\mu_1.$$

Indeed, we can verify that

$$\begin{aligned} &\int_{a_1}^{b_1} y^k R_{n_2}(y) \frac{P_{n_1,n_2}(y) f_{1,2}(y) - Q_{n_1,n_2;1}(y)}{R_{n_2}(y)} \, d\mu_1(y) \\ &\qquad = \int_{a_1}^{b_1} y^k [P_{n_1,n_2}(y) f_{1,2}(y) - Q_{n_1,n_2;1}(y)] \, d\mu_1(y) = 0, \end{aligned}$$

for $k \leq n_2 - 1$ since we have seen that the expression $P_{n_1,n_2} f_{1,2} - Q_{n_1,n_2;1}$ is orthogonal to all polynomials of degree less than n_2 on $[a_1, b_1]$ for the measure μ_1. The orthogonality measure for R_{n_2} can also be written as

$$\frac{P_{n_1,n_2}(y) f_{1,2}(y) - Q_{n_1,n_2;1}(y)}{R_{n_2}(y)} = \frac{1}{P_{n_1,n_2}(y)} \int_{a_2}^{b_2} \frac{P_{n_1,n_2}^2(x)}{y - x} \frac{d\mu_2(x)}{R_{n_2}(x)}.$$

In this weight we have

$$\frac{1}{\gamma_{n_1,n_2}^2 C_1} \leq \int_{a_2}^{b_2} \frac{P_{n_1,n_2}^2(x)}{|y - x|} \frac{d\mu_2(x)}{|R_{n_2}(x)|} \leq \frac{1}{\gamma_{n_1,n_2}^2 C_2},$$

where C_1 and C_2 are the maximum and minimum, respectively, over the set

$$\{|x - y| : x \in [a_2, b_2], \ y \in [a_1, b_1]\}.$$

So, up to the constants C_1, C_2, we have the extremal problem

$$\begin{aligned} \frac{1}{\gamma_{n_2;2}^2} &= \int_{a_1}^{b_1} R_{n_2}^2(y) \frac{d\mu_1(y)}{|P_{n_1,n_2}(y)|} \\ &= \min_{\pi_{n_2}(y)=y^{n_2}+\cdots} \int_{a_1}^{b_1} \pi_{n_2}^2(y) \frac{d\mu_1(y)}{|P_{n_1,n_2}(y)|}. \end{aligned} \tag{2.21}$$

Define the zero distributions

$$\nu_{n_1+n_2} := \frac{1}{n_1+n_2} \sum_{j=1}^{n_1+n_2} \delta_{x_{j,n_1+n_2}}, \quad \nu_{n_2;2} := \frac{1}{n_2} \sum_{j=1}^{n_2} \delta_{y_{j,n_2}},$$

where x_{j,n_1+n_2} are the zeros of P_{n_1,n_2} and y_{j,n_2} are the zeros of R_{n_2}. Then $(\nu_{n_1+n_2})$ is a sequence of probability measures on $[a_2, b_2]$ and $(\nu_{n_2;2})$ is a sequence of probability measures on $[a_1, b_1]$. Helly's selection principle shows that there are weakly convergent subsequences with limits ν and ν_2 which are supported on $[a_2, b_2]$ and $[a_1, b_1]$ respectively. The extremal problems (2.20) and (2.21) then lead to an extremal problem in potential theory. The integral in (2.20) is approximately

$$\int_{a_2}^{b_2} \exp[-2(n_1+n_2)U(x;\nu) + n_2 U(x;\nu_2)]\, d\mu_2(x)$$

and the integral in (2.21) is approximately

$$\int_{a_1}^{b_1} \exp[-2n_2 U(x;\nu_2) + (n_1+n_2)U(x;\nu)]\, d\mu_1(x).$$

If $n_2/(n_1+n_2) \to q$ and $\mu_i' > 0$ almost everywhere on $[a_i, b_i]$ $(i = 1, 2)$, then this gives the variational conditions

$$\begin{aligned} 2U(x;\nu) - qU(x;\nu_2) &= \ell_1, \qquad x \in \operatorname{supp}(\nu) \subset [a_2, b_2], && (2.22)\\ -U(x;\nu) + 2qU(x;\nu_2) &= \ell_2, \qquad x \in \operatorname{supp}(\nu_2) \subset [a_1, b_1], && (2.23) \end{aligned}$$

where ℓ_1 and ℓ_2 are Lagrange multipliers for which

$$\lim_{n_1+n_2\to\infty} \gamma_{n_1,n_2}^{2/(n_1+n_2)} = \exp(\ell_1), \qquad \lim_{n_1+n_2\to\infty} \gamma_{n_2;2}^{2/(n_1+n_2)} = \exp(\ell_2).$$

Looking back to (2.19) we thus have

$$\lim_{n_1+n_2\to\infty} \left| f_{1,2}(y) - \frac{Q_{n_1,n_2;1}(y)}{P_{n_1,n_2}(y)} \right|^{1/(n_1+n_2)} = r < 1$$

on level curves $C_r := \{z : \exp[2U(z;\nu) - qU(z;\nu_2) - \ell_1] = r\}$.

The convergence to the second function $f_{2,2}$ can also be handled but is left as an advanced exercise for the reader.

3 Applications

3.1 Gauss and simultaneous Gauss quadrature

Gauss quadrature is directly related to orthogonal polynomials, and hence to Padé approximation. Here is an approach based on complex analysis. Suppose μ is a positive measure on $[a, b]$ and we denote by f the Markov function for μ,

$$f(z) = \int_a^b \frac{d\mu(x)}{z-x}.$$

Let Q_{n-1}/P_n be the Padé approximant to f near infinity. Then

$$f(z) - \frac{Q_{n-1}(z)}{P_n(z)} = \mathcal{O}(z^{-2n-1}), \qquad z \to \infty.$$

Multiply both sides by a polynomial π_{2n-1} of degree at most $2n-1$, and integrate along a contour Γ encircling the interval $[a,b]$ once in the positive direction. Then

$$\frac{1}{2\pi i}\int_\Gamma \pi_{2n-1}(z) f(z)\,dz = \frac{1}{2\pi i}\int_\Gamma \pi_{2n-1}(z)\frac{Q_{n-1}(z)}{P_n(z)}\,dz,$$

because the remainder term vanishes after integration, due to Cauchy's theorem for the outside of Γ. Interchanging the order of integration on the left hand side and using the residue theorem on the right hand side shows that for every polynomial π_{2n-1} of degree $\leq 2n-1$ we have

$$\int_a^b \pi_{2n-1}(x)\,d\mu(x) = \sum_{j=1}^n \lambda_{j,n}\pi_{2n-1}(x_{j,n}), \tag{3.1}$$

where $\lambda_{j,n}$ is the residue of the Padé approximant at the zeros $x_{j,n}$ of P_n, i.e.,

$$\lambda_{j,n} = \frac{Q_{n-1}(x_{j,n})}{P_n'(x_{j,n})}.$$

If we take $\pi_{2n-1}(x) = P_n^2(x)/(x-x_{j,n})^2$, then (3.1) gives

$$\int_a^b \frac{P_n^2(x)}{(x-x_{j,n})^2}\,d\mu(x) = \lambda_{j,n}[P_n'(x_{j,n})]^2,$$

which shows that $\lambda_{j,n} > 0$ for $j = 1,\ldots,n$. These weights $\lambda_{j,n}$ are known as Christoffel numbers or Gauss quadrature coefficients, the zeros $x_{j,n}$ of P_n are Gauss quadrature nodes, and (3.1) is the Gauss quadrature formula. Replacing π_{2n-1} by a continuous function g on $[a,b]$ suggests to use the sum

$$\sum_{j=1}^n \lambda_{j,n} g(x_{j,n})$$

as an approximation to the integral

$$\int_a^b g(x)\,d\mu(x).$$

If $[a,b]$ is a finite interval, then every continuous function can be approximated uniformly by polynomials (Weierstrass), hence the quadrature sum indeed converges to the integral when the number of nodes n tends to infinity. The positivity of the weights $\lambda_{j,n}$ is needed to get this convergence. The quadrature formula requires n function evaluations (at the zeros of P_n) and is exact for polynomials of degree $\leq 2n-1$, hence on a linear space of dimension $2n$. The ratio $n/2n = 1/2$ is a measure for the efficiency of this formula.

In a number of applications, we need to approximate several integrals of the same function but with respect to different measures. The following example comes from [9]. Suppose that g is the spectral distribution of light in the direction of the observer and w_1, w_2, w_3 are weight functions describing the profiles for red, green and blue light. Then the integrals

$$\int_0^{2\pi} g(x)w_1(x)\,dx, \quad \int_0^{2\pi} g(x)w_2(x)\,dx, \quad \int_0^{2\pi} g(x)w_3(x)\,dx$$

give the amount of light after passing through the filters for red, green and blue. In this case, we need to approximate three integrals of the same function g. We would like to use as few function evaluations as possible, but the integrals should be accurate for polynomials g of degree as high as possible. If we use Gauss quadrature with n nodes for each integral, then we require $3n$ function evaluations and all integrals will be correct for polynomials of degree $\leq 2n-1$ (a space of dimension $2n$). This gives an efficiency of 3/2. In fact, with $3n$ function evaluations, we can double the dimension of the space in which the formula is exact. Consider the Markov functions

$$f_j(z) = \int_0^{2\pi} \frac{w_j(x)\,dx}{z-x}, \qquad j = 1, 2, 3$$

and the type II Hermite-Padé approximation problem

$$f_j(z) - \frac{Q_{n,n,n;j}(z)}{P_{n,n,n}(z)} = \mathcal{O}(z^{-4n-1}), \qquad z \to \infty.$$

Now we can multiply by a polynomial π_{4n-1} of degree at most $4n-1$, and integrate along a contour Γ encircling $[0, 2\pi]$ in the positive direction, to obtain

$$\int_0^{2\pi} \pi_{4n-1}(x)w_j(x)\,dx = \sum_{k=1}^{3n} \lambda_{k,n;j}\pi_{4n-1}(x_{k,n}), \qquad j = 1, 2, 3, \tag{3.2}$$

where $x_{k,n}$ are the zeros of $P_{n,n,n}$ and $\lambda_{k,n;j}$ are the residues of $Q_{n,n,n;j}/P_{n,n,n}$ at the zero $x_{k,n}$:

$$\lambda_{k,n;j} = \frac{Q_{n,n,n;j}(x_{k,n})}{P'_{n,n,n}(x_{k,n})}.$$

Therefore, the three integrals will be evaluated exactly by the three sums in (3.2) for polynomials of degree $\leq 4n-1$. The convergence is somewhat more difficult to handle since we do not have a general result that the quadrature coefficients $\lambda_{k,n;j}$ are positive. The positivity has to be investigated separately for Angelesco and Nikishin systems. See [13, 16, 17] for finding out more about simultaneous Gauss quadrature.

3.2 Irrationality and transcendence

Hermite-Padé approximants were introduced by Hermite in his proof that e is transcendental. Various irrationality proofs of famous mathematical constants use Hermite-Padé approximation, even though this may not always be obvious. Proving irrationality can be done by constructing good rational approximants:

Lemma 3.1. *Let $x \in \mathbb{R}$. Suppose we can find sequences of integers $(p_n), (q_n)$ such that*

1. *$q_n x - p_n \neq 0$ for all $n \in \mathbb{N}$,*

2. $\lim_{n\to\infty}(q_n x - p_n) = 0$.

Then x is irrational.

Proof. Suppose that x is rational. Then $x = p/q$ for some integers p, q. We then have

$$q_n x - p_n = \frac{q_n p - p_n q}{q}$$

and since this is not zero for every n, we see that $q_n p - p_n q \neq 0$ for all n. But since these are integers, this implies that $|q_n p - p_n q| \geq 1$ for all n. This shows that $|q_n x - p_n| \geq 1/q$, which is in contradiction with condition 2 in the lemma. Hence we must conclude that x is irrational. ☐

The construction of the sequences (p_n) and (q_n) often uses Padé or Hermite-Padé approximation for well chosen functions. As an example, consider the two Markov functions

$$f_1(z) := \int_0^1 \frac{dx}{z-x}, \quad f_2(z) := \int_{-1}^0 \frac{dx}{z-x},$$

which form an Angelesco system. Some straightforward calculus gives

$$f_1(i) = -\frac{1}{2}\log 2 - \frac{i\pi}{4}, \quad f_2(i) = \frac{1}{2}\log 2 - \frac{i\pi}{4},$$

hence the sum gives $f_1(i) + f_2(i) = -i\pi/2$. The type II Hermite-Padé approximants for f_1 and f_2 will give approximations to π. Recall that

$$\begin{aligned} P_{n,n}(z)f_1(z) - Q_{n,n;1}(z) &= \int_0^1, \frac{P_{n,n}(x)}{z-x}\,dx \\ P_{n,n}(z)f_2(z) - Q_{n,n;2}(z) &= \int_{-1}^0 \frac{P_{n,n}(x)}{z-x}\,dx. \end{aligned}$$

Summing both equations gives

$$P_{n,n}(z)[f_1(z) + f_2(z)] - [Q_{n,n;1}(z) + Q_{n,n;2}(z)] = \int_{-1}^1 \frac{P_{n,n}(x)}{z-x}\,dx.$$

So, the fact that we are using a common denominator comes in very handy here. Then we evaluate these expressions at $z = i$ and hope that $P_{n,n}(i)$ and $Q_{n,n;1}(i) + Q_{n,n;2}(i)$ are (up to the factor i) integers or rational numbers with simple denominators. Conditions 1 and 2 in Lemma 3.1 can be checked by using asymptotic properties of Hermite-Padé approximation. For this particular case, the type II multiple orthogonal polynomials are given by a Rodrigues formula

$$P_{n,n}(x) = \frac{d^n}{dx^n}\left(x^n(1-x^2)^n\right),$$

and these polynomials are known as **Legendre-Angelesco polynomials**. They have been studied in detail by Kalyagin [22] (see also [32]). The Rodrigues formula in fact simplifies the asymptotic analysis since integration by parts now gives

$$\int_{-1}^1 \frac{P_{n,n}(x)}{z-x}\,dx = \int_{-1}^1 (-1)^n n! \frac{x^n(1-x^2)^n}{(z-x)^{n+1}}\,dx,$$

which can be handled easily. Some trial and error show that one gets better results by taking $2n$ instead of n, and by differentiating n times:

$$\frac{d^n}{dz^n}\left(P_{2n,2n}(z)[f_1(z)+f_2(z)]-[Q_{2n,2n;1}(z)+Q_{2n,2n;2}(z)]\right)_{z=i} = (3n)!(-i)^{n+1}\int_{-1}^{1}\frac{x^{2n}(1-x^2)^{2n}}{(1+ix)^{3n+1}}\,dx. \quad (3.3)$$

This gives rational approximants to π of the form

$$\pi = \frac{b_n}{a_n c_n} + \frac{K_n}{a_n},$$

where a_n, b_n, c_n are explicitly known integers and K_n is the integral on the right hand side of (3.3). The rational approximants show that π is irrational (which was shown already in 1773 by Lambert), and they even show that you can't approximate π by rationals at order greater than 23.271 (Beukers [6]), i.e.,

$$\left|\pi - \frac{p}{q}\right| < \frac{1}{q^r},$$

with $r > 23.271$, only has a finite number of solutions (p,q), where p and q are relatively prime integers. This upper bound for the order of approximation can be reduced to 8.02 (Hata [20]) by considering Markov functions f_1 and f_3, with

$$f_3(z) := \int_{-i}^{0}\frac{dx}{z-x}.$$

This f_3 is now over a complex interval, and then Theorem 2.1 concerning the location of the zeros no longer holds, and the asymptotic behavior must be handled by another method.

One can also use Hermite-Padé approximants to prove transcendence. Then one uses the following lemma which extends Lemma 3.1 from irrational numbers to non-algebraic numbers.

Lemma 3.2. *Let $x \in \mathbb{R}$. Suppose that, for every integer $m \in \mathbb{N}$ and for all integers $a_0, a_1, \ldots, a_m \in \mathbb{Z}$, we can find integers $p_{0,n}, p_{1,n}, \ldots, p_{m,n}$ such that*

1. *$\sum_{k=0}^{m} a_k p_{k,n} \neq 0$ for all $n \in \mathbb{N}$,*
2. *$\lim_{n\to\infty}(p_{0,n}x^k - p_{k,n}) = 0$ for $k = 1, 2, \ldots, m$.*

Then x is transcendental.

Proof. Suppose that x is algebraic. Then there exists an integer m and integers $a_0, \ldots, a_m$ such that $\sum_{k=0}^{m} a_k x^k = 0$. But then

$$\sum_{k=0}^{m} a_k(p_{0,n}x^k - p_{k,n}) = -\sum_{k=0}^{m} a_k p_{k,n}.$$

The right hand side is an integer different from zero, hence

$$\left|\sum_{k=0}^{m} a_k(p_{0,n}x^k - p_{k,n})\right| \geq 1, \quad \text{for all } n \in \mathbb{N}.$$

But this contradicts condition 2 of the lemma. Hence we must conclude that x is not algebraic. $\square$

If we use type II Hermite-Padé approximation to $(e^{\lambda_1 x}, e^{\lambda_2 x}, \ldots, e^{\lambda_r x})$ near $x = 0$, then this will give the transcendence of e. For Hermite-Padé approximation near $x = 0$ we can use two multi-indices $\vec{n} = (n_1, n_2, \ldots, n_r)$ and $\vec{m} = (m_1, m_2, \ldots, m_r)$. These Hermite-Padé approximants are known explicitly when $m_j + n_j = N + |\vec{n}|$ for $1 \leq j \leq r$, where N is an integer. If we define the polynomial

$$T(x) := x^N (x - \lambda_1)^{n_1} (x - \lambda_2)^{n_2} \cdots (x - \lambda_r)^{n_r},$$

then T has degree $N + |\vec{n}|$. The expression

$$P_{\vec{n}}(z) = z^{|\vec{n}|+N+1} \int_0^\infty T(x) e^{-zx}\, dx$$

gives a polynomial of degree $|\vec{n}|$, and

$$Q_{\vec{m};j}(z) = z^{|\vec{n}|+N+1} \int_0^\infty T(x + \lambda_j) e^{-zx}\, dx$$

gives a polynomial of degree $|\vec{n}| + N - n_j = m_j$. One easily verifies that

$$P_{\vec{n}}(z) e^{\lambda_j z} - Q_{\vec{m};j}(z) = e^{\lambda_j z} z^{|\vec{n}|+N+1} \int_0^{\lambda_j} T(x) e^{-zx}\, dx = \mathcal{O}(z^{n_j + m_j + 1}),$$

as $z \to 0$, which are the interpolation conditions for type II Hermite-Padé approximation near the origin for the two multi-indices $(\vec{n}, \vec{m})$.

For proving the transcendence of e, we take $\lambda_j = j$, $z = 1$ and for a prime $p > r$, which is not a divisor of a_0, we take $N = p - 1$ and $n_j = p$ $(j = 1, \ldots, r)$. Then some elementary calculus shows that $p_0 = P_{\vec{n}}(1)/(p-1)!$ is an integer which is not divisible by p and each $p_j = Q_{\vec{m};j}(1)/(p-1)!$ is an integer divisible by p. Therefore $\sum_{j=0}^r a_j p_j$ is not divisible by p and hence condition 1 of Lemma 3.2 is satisfied. Furthermore

$$p_0 e^j - p_j = \frac{e^j}{(p-1)!} \int_0^j T(x) e^{-x}\, dx,$$

and the simple estimate $|T(x)| \leq j^{(r+1)p-1}$ on $[0, j]$ shows that this converges to 0 for every j when the prime p tends to infinity (luckily Euclides showed that there are infinitely many primes). So condition 2 of Lemma 3.2 is also satisfied and we conclude that e is transcendental (Hermite, 1874).

3.3 Other applications

Recently, a number of applications came up in other areas of mathematics and theoretical physics. There are interesting connections with random matrix theory, where multiple orthogonal polynomials (in particular multiple Hermite polynomials) appear when one investigates random matrices with an external source [8, 5]. Multiple Laguerre polynomials appear for the Wishart ensemble of random matrices [7]. Multiple Jacobi polynomials (the Jacobi-Piñeiro polynomials) were used to obtain a counterexample to the Bethe Ansatz Conjecture for the Gaudin model [25]. More details on multiple orthogonal polynomials (recursion relation, specific examples, etc.) can be found in [21, Chapter 23].

Acknowledgments

Research supported by research grant OT/04/21 of Katholieke Universiteit Leuven, research project G.0455.04 of FWO-Vlaanderen and INTAS research network 03-51-6637.

References

[1] M. A. Angelesco, *Sur deux extensions des fractions continues algébriques*, C.R. Acad. Sci. Paris **18** (1919), 262–263.

[2] A. I. Aptekarev, *Asymptotics of simultaneously orthogonal polynomials in the Angelesco case*, Mat. Sb. **136** (178) (1988), 56–84 (Russian); Math. USSR Sb. **64** (1989), 57–84.

[3] A. I. Aptekarev, *Multiple orthogonal polynomials*, J. Comput. Appl. Math. **99** (1998), 423–447.

[4] A. I. Aptekarev, *Strong asymptotics of multiple orthogonal polynomials for Nikishin systems*, Mat. Sb. **190** no. 5 (1999), 3–44 (Russian); Sbornik Math. **190** no. 5 (1999), 631–669.

[5] A. I. Aptekarev, P. M. Bleher, A. B. J. Kuijlaars, *Large n limit of Gaussian random matrices with external source. II*, Comm. Math. Phys. **259** (2005), no. 2, 367–389.

[6] F. Beukers, *A rational approach to* π, Nieuw Arch. Wisk. (5) **1** (2000), no. 4, 372–379.

[7] P. M. Bleher, A. B. J. Kuijlaars, *Integral representations for multiple Hermite and multiple Laguerre polynomials* Ann. Inst. Fourier (Grenoble) **55** (2005), no. 6, 2001–2014.

[8] P. Bleher, A.B.J. Kuijlaars, *Large n limit of Gaussian random matrices with external source. I*, Comm. Math. Phys. **252** (2004), no. 1–3, 43–76.

[9] C. F. Borges, *On a class of Gauss-like quadrature rules*, Numer. Math. **67** (1994), 271–288.

[10] M. G. de Bruin, *Simultaneous Padé approximation and orthogonality*, in "Polynômes Orthogonaux et Applications" (C. Brezinski et al., eds.), Lecture Notes in Mathematics **1171**, Springer-Verlag, Berin, 1985, pp. 74–83.

[11] M. G. de Bruin, *Some aspects of simultaneous rational approximation*, in "Numerical Analysis and Mathematical Modelling", Banach Center Publications **24**, PWN-Polish Scientific Publishers, Warsaw, 1990, pp. 51–84.

[12] J. Bustamante, G. López Lagomasino, *Hermite-Padé approximation to a Nikishin type system of analytic functions*, Mat. Sb. **183** (1992), 117–138 (Russian); Acad. Sci. Sb. Math. **77** (1994), 367–384.

[13] J. Coussement, W. Van Assche, *Gaussian quadrature for multiple orthogonal polynomials*, J. Comput. Appl. Math. **178** (2005), 131–145.

[14] K. Driver, H. Stahl, *Normality in Nikishin systems*, Indag. Math. (New Series) **5** (1994), 161–187.

[15] K. Driver, H. Stahl, *Simultaneous rational approximants to Nikishin systems. I-II*, Acta Sci. Math. (Szeged) **60** (1995), 245–263; **61** (1995), 261–284.

[16] U. Fidalgo Prieto, J. Illán, G. López Lagomasino, *Hermite-Padé approximation and simultaneous quadrature formulas*, J. Approx. Theory **126** (2004), 171–197.

[17] U. Fidalgo Prieto, J.R. Illán González, G. López Lagomasino, *Convergence and computation of simultaneous rational quadrature formulas*, Numer. Math. (to appear)

[18] A. A. Gonchar, E. A. Rakhmanov, *On the equilibrium problem for vector potentials*, Usp. Mat. Nauk **40** no. 4 (244) (1985), 155–156 (Russian); Russ. Math. Surveys **40** no. 4 (1985), 183–184.

[19] A. A. Gonchar, E. A. Rakhmanov, V. N. Sorokin, *Hermite-Padé approximants for systems of Markov-type functions*, Mat. Sb. **188** (1997), 38–58 (Russian); Russian Acad. Sb. Math. **188** (1997), 671–696.

[20] M. Hata, *Rational approximations to π and some other numbers*, Acta Arith. **63** (1993), no. 4, 335–349.

[21] M. E. H. Ismail, *Classical and Quantum Orthogonal Polynomials in One Variable*, Encyclopedia of Mathematics and its Applications vol. 98, Cambridge University Press, 2005.

[22] V. A. Kalyagin (Kaliaguine), *On a class of polynomials defined by two orthogonality conditions*, Mat. Sb. **110** (1979), 609–627 (Russian); Math. USSR Sb. **38** (1981), 563–580.

[23] S. Karlin, W. J. Studden, Tchebycheff Systems: with Applications in Analysis and Statistics, Pure and Applied Mathematics, Vol. XV, Interscience Publishers, John Wiley & Sons, New York, 1966.

[24] K. Mahler, *Perfect systems*, Compos. Math. **19** (1968), 95–166.

[25] E. Mukhin, A. Varchenko, *Multiple orthogonal polynomials and a counterexample to Gaudin Bethe Ansatz conjecture*, Trans. Amer. Math. Soc. (to appear), `arXiv.org:math/0501144`

[26] E. M. Nikishin, *On simultaneous Padé approximations*, Mat. Sb. **113** (155) (1980), 499–519 (Russian); Math. USSR Sb. **41** (1982), 409–425.

[27] E. M. Nikishin, V. N. Sorokin, Rational Approximations and Orthogonality, Translations of Mathematical Monographs vol. 92, Amer. Math. Soc., Providence, RI, 1991.

[28] J. Nuttall, *Asymptotics of diagonal Hermite-Padé polynomials*, J. Approx. Theory **42** (1984), 299–386.

[29] H. Padé, *Mémoire sur les développements en fractions continues de la fonction exponentielle pouvant servir d'introduction à la théorie des fractions continues algébriques*, Ann. Sci. Ecole Norm. Sup. (3) **16** (1899), 395–426.

[30] E. B. Saff, V. Totik, Logarithmic Potentials with External Fields, Grundlehren der Mathematischen Wissenschaften **136**, Springer-Verlag, Berlin, 1997.

[31] H. Stahl, V. Totik, General Orthogonal Polynomials, Encyclopedia of Mathematics and its Applications vol. 43, Cambridge University Press, 1992.

[32] W. Van Assche, *Multiple orthogonal polynomials, irrationality and transcendence*, in "Continued fractions: from analytic number theory to constructive approximation" (Columbia, MO, 1998), Contemporary Mathematics **236** (1999), 325–342.

[33] W. Van Assche, E. Coussement, *Some classical multiple orthogonal polynomials*, in "Numerical analysis 2000, Vol. V, Quadrature and Orthogonal Polynomials", J. Comput. Appl. Math. **127** (2001), no. 1-2, 317–347.

Walter Van Assche
Katholieke Universiteit Leuven
Department of Mathematics
Celestijnenlaan 200B
B-3001 Leuven
BELGIUM
walter@wis.kuleuven.be
http://www.wis.kuleuven.be/analyse/walter/