Approximationstheorie

Ergänzungen

Oliver Ernst

Professur Numerische Mathematik

Sommersemester 2018

Mathematik!
TU Chemnitz

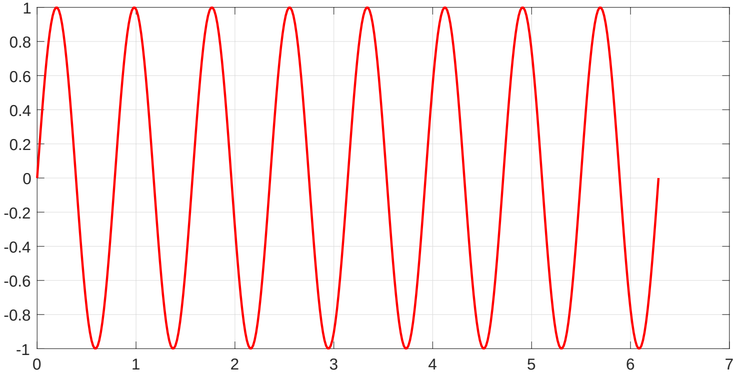
Contents I

Contents

Aliasing

Most familiar setting

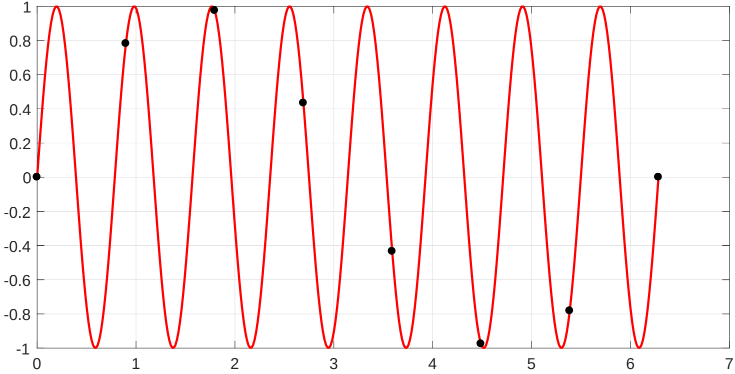
$f(x) = \sin(8x)$



Aliasing

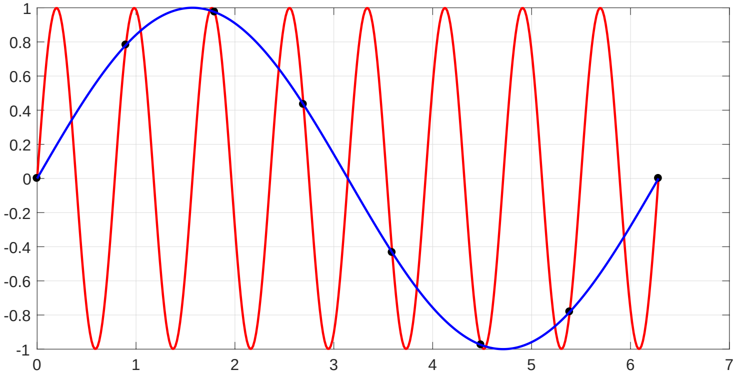
Most familiar setting

$f(x) = \sin(8x)$, sampled at $x_j = j \cdot 2\pi/7$



Aliasing
Most familiar setting

$f(x) = \sin(8x)$, sampled at $x_j = j \cdot 2\pi/7$



aliases to $\tilde{f}(x) = \sin x$.

Aliasing

Chebyshev, Laurent and Fourier

Recall that a Lipschitz continuous function f on $[-1,1]$ has the absolutely and uniformly convergent Chebyshev expansion $f(x) = \sum_{k=0}^{\infty} a_k T_k(x)$.
Given $n \in \mathbb{N}_0$, we also recall the Chebyshev projection and Chebyshev interpolant

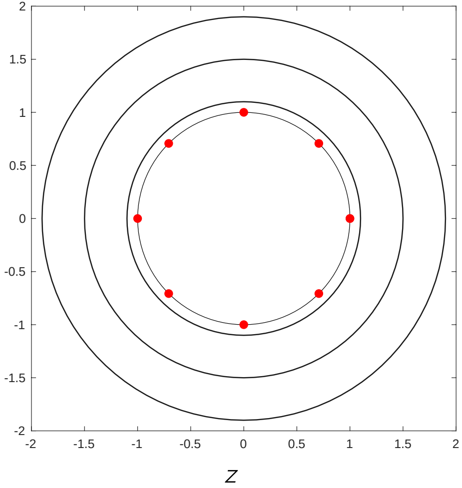
$$p_n(x) = \sum_{k=0}^{n} c_k T_k(x), \qquad f_n(x) = \sum_{k=0}^{n} a_n T_k(x), \qquad x \in [-1,1].$$

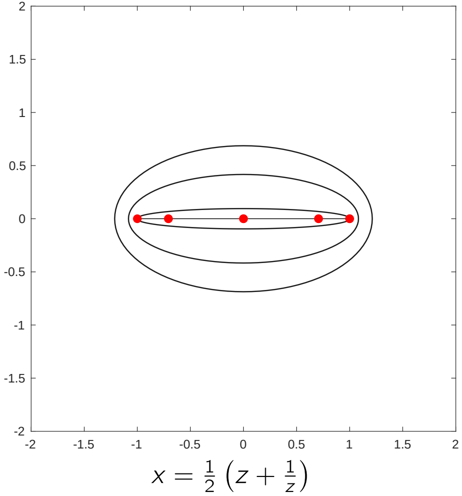
In the variables z and ϑ, where $x = \frac{1}{2}(z + z^{-1})$ and $z = \mathrm{e}^{i\vartheta}$, the corresponding interpolations and projections are as follows:

Fourier	**Laurent**
$\mathbb{F}(\vartheta) = \mathbb{F}(-\vartheta) = \frac{1}{2}\sum_{k=0}^{\infty} a_k \left(\mathrm{e}^{ik\vartheta} + \mathrm{e}^{-ik\vartheta}\right)$	$F(z) = F(z^{-1}) = \frac{1}{2}\sum_{k=0}^{\infty} a_k \left(z^k + z^{-k}\right)$
$\mathbb{F}_n(\vartheta) = \frac{1}{2}\sum_{k=0}^{n} a_k \left(\mathrm{e}^{ik\vartheta} + \mathrm{e}^{-ik\vartheta}\right)$	$F_n(z) = \frac{1}{2}\sum_{k=0}^{n} a_k \left(z^k + z^{-k}\right)$
$\mathbb{P}(\vartheta) = \frac{1}{2}\sum_{k=0}^{n} c_k \left(\mathrm{e}^{ik\vartheta} + \mathrm{e}^{-ik\vartheta}\right)$	$P_n(z) = \frac{1}{2}\sum_{k=0}^{n} c_k \left(z^k + z^{-k}\right)$

Aliasing
Joukowsky map

Image of circles around the origin.

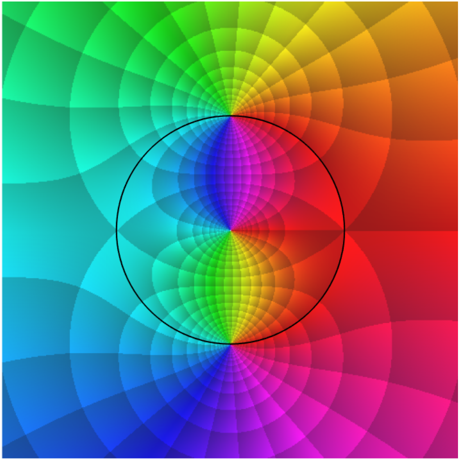




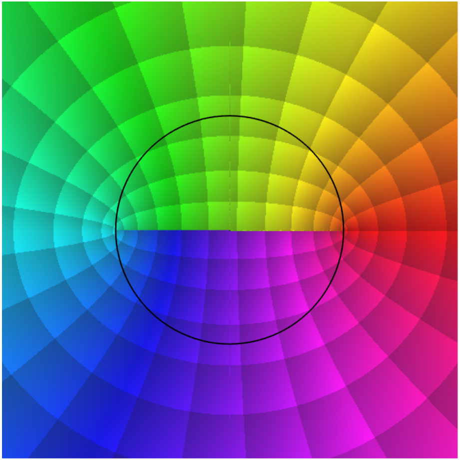
Aliasing

Joukowsky map

Phase plots.



z



$x = \frac{1}{2}\left(z + \frac{1}{z}\right)$

Contents

Barycentric Interpolation Formula

Review

We recall some basic facts about Lagrange interpolation:

- For a set of $n+1$ distinct interpolation nodes $\{x_j\}_{j=0}^n$ and $n+1$ data values $\{f_j\}_{j=0}^n$ there exists a unique polynomial $p \in \mathscr{P}_n$ satisfying the interpolation conditions

$$p(x_j) = f_j, \qquad j = 0, \dots, n.$$

- The interpolant may be represented in terms of the Lagrange fundamental polynomials $\{\ell_j\}_{j=0}^n \subset \mathscr{P}_n$ as

$$p(x) = \sum_{j=0}^{n} f_j \, \ell_j(x),$$

where

$$\ell_j(x) = \prod_{\substack{k=0 \\ k \neq j}}^{n} \frac{x - x_k}{x_j - x_k}, \qquad j = 0, \dots, n.$$

Barycentric Interpolation Formula

Review

- The Newton form of the interpolating polynomial is based on the nodal polynomials[1]

$$\omega_j(x) = \prod_{k=0}^{j-1} (x - x_k) \in \mathscr{P}_j, \qquad j = 0, \dots, n,$$

as well as the set of recursively defined divided differences

$$f_{i_0,i_1,\dots,i_k} := \frac{f_{i_1,i_2,\dots,i_k} - f_{i_0,i_1,\dots,i_{k-1}}}{x_{i_k} - x_{i_0}} \qquad k \geq 1,$$

where $i_0, \dots, i_n \in \{0, 1, \dots, n\}$ are distinct indices, and has the form

$$p(x) = f_0\, \omega_0(x) + f_{0,1}\, \omega_1(x) + \cdots + f_{0,1,\dots,n}\, \omega_n(x).$$

[1] (the empty product taken as equal to one)

- The divided differences can be generated from left to right in the triangular table:

x_i	$k=0$	$k=1$	$k=2$	$k=3$	$k=4$
x_0	f_0				
		$f_{0,1}$			
x_1	f_1		$f_{0,1,2}$		
		$f_{1,2}$		$f_{0,1,2,3}$	
x_2	f_2		$f_{1,2,3}$		$f_{0,1,2,3,4}$
		$f_{2,3}$		$f_{1,2,3,4}$	
x_3	f_3		$f_{2,3,4}$		
		$f_{3,4}$			
x_4	f_4				

Adding a data pair entails adding a ‘diagonal’ along the bottom of the triangle.

The representation of the Lagrange interpolating polynomial with respect to the monomial basis $\{x^j\}_{j=0}^n$, i.e.,

$$p(x) = a_0 + a_1 x + \cdots + a_n x^n$$

is determined by the $n+1$ linear equations

$$p(x_j) = f_j, \qquad j = 0, \dots, n,$$

or, in matrix form, $\mathbf{Va} = \mathbf{f}$, where

$$\mathbf{V} = \mathbf{V}(x_0, \dots, x_n) = \begin{bmatrix} 1 & x_0 & x_0^2 & \dots & x_0^n \\ 1 & x_1 & x_1^2 & \dots & x_1^n, \\ \vdots & & \vdots & & \\ 1 & x_n & x_n^2 & \dots & x_n^n \end{bmatrix}, \quad \mathbf{a} = \begin{bmatrix} a_0 \\ a_1 \\ \vdots \\ a_n \end{bmatrix}, \quad \mathbf{f} = \begin{bmatrix} f_0 \\ f_1 \\ \vdots \\ f_n \end{bmatrix}.$$

In view of $\det \mathbf{V}(x_0, \dots, x_n) = \prod_{i>j}(x_i - x_j)$, this has a unique solution whenever the nodes are distinct.

In terms of the nodal polynomial associated with $\{x_j\}_{j=0}^n$

$$\omega_{n+1}(x) := (x - x_0)(x - x_1)\cdots(x - x_n) \in \mathscr{P}_{n+1}.$$

we define the barycentric weights $\{\lambda_j\}_{j=0}^n$ by

$$\lambda_j := \frac{1}{\prod_{\substack{k=0\\k\neq j}}^n (x_j - x_k)} = \frac{1}{\omega'_{n+1}(x_j)}, \qquad j = 0, \dots, n, \tag{5.1}$$

in terms of which the Lagrange fundamental polynomials become

$$\ell_j(x) = \omega_{n+1}(x)\frac{\lambda_j}{x - x_j}, \qquad j = 0, \dots, n,$$

by means of which we can represent the interpolating polynomial as ...

Barycentric Interpolation Formula

First and second formulas

the first barycentric formula (or modified Lagrange formula)

$$p(x) = \omega_{n+1}(x) \sum_{j=0}^{n} f_j \frac{\lambda_j}{x - x_j}.$$

Since the constant function $f \equiv 1$ is always interpolated exactly, we have

$$1 \equiv \omega_{n+1}(x) \sum_{j=0}^{n} \frac{\lambda_j}{x - x_j},$$

so that, dividing the first barycentric formula by this expression and cancelling common factors yields the second barycentric formula

$$p(x) = \frac{\displaystyle\sum_{j=0}^{n} f_j \frac{\lambda_j}{x - x_j}}{\displaystyle\sum_{j=0}^{n} \frac{\lambda_j}{x - x_j}}.$$

Updating. Addition of new node x_{n+1}:

$$\lambda_j^{\text{new}} := \frac{\lambda_j^{\text{old}}}{x_j - x_{n+1}}, \qquad j = 0, \dots, n, \quad (2n+2 \text{ flops}).$$

λ_{n+1} from old weights, additional $n+1$ flops, if $x_j - x_{n+1}$ have been stored.

Cost.

- Computation of $\{\lambda_j\}_{j=0}^n$ requires $\sum_{j=1}^n 3j = \frac{3}{2}n(n+1)$ flops.
- For given weights $\{\lambda_j\}_{j=0}^n$ each evaluation of p in additional $5n+4 = O(n)$ flops.

Further advantages.

- λ_j independent of f_j, i.e., once weights computed arbitrary f can be interpolated in $O(n)$ flops.
- λ_j independent of node numbering (cf. divided differences).

Theorem 5.1

For the $n+1$ Chebyshev nodes the barycentric weights are

$$\lambda_j = \begin{cases} (-1)^j \frac{2^{n-1}}{n}, & j = 1, \dots, n-1, \\ (-1)^j \frac{2^{n-2}}{n}, & j = 0, n. \end{cases}$$

The associated interpolating polynomial for data $\{f_j\}_{j=0}^n$ in second barycentric form is then

$$p(x) = \frac{\displaystyle\sum_{j=0}^{n}{}' \frac{(-1)^j f_j}{x - x_j}}{\displaystyle\sum_{j=0}^{n}{}' \frac{(-1)^j}{x - x_j}},$$

with the primed sums indicating that the first and last term are halved.

Barycentric Interpolation Formula

Remarks

- The exponential growth of the barycentric weights for Chebyshev nodes raises concern about floating point overflow for high interpolation degrees.
- Moreover, the nodal polynomial occurring as a factor in the first barycentric formula has value on the order of 2^{-n} on $[-1, 1]$, which similarly poses a danger of underflow.
- In addition, all these numbers scale with n-th powers when the independent variable is transplanted to a general bounded interval $[a, b] \subset \mathbb{R}$.
- The over- and underflow issues can be addressed by reformulating the expressions in terms of logarithms or mapping the independent variable to an interval of length 4 (logarithmic capacity 1).
- The nodal polynomial factor and the common factors in the barycentric weight, however, cancel out in the second barycentric formula, making it scale invariant and eliminating the risk of over- and underflow.
- Nonetheless, the second barycentric formula has weaker numerical stability properties than the first, as the analysis in [Higham, 2004] makes explicit.

Definition 5.2

The condition number of $p = p_n$ at $x \in [-1, 1]$ with respect to f is, for $p(x) \neq 0$,

$$\mathrm{cond}(x, n, f) := \lim_{\epsilon \to 0} \sup \left\{ \left| \frac{p_f(x) - p_{f+\Delta f}(x)}{\epsilon p_f(x)} \right| : |\Delta f| \leq \epsilon |f| \right\}$$

In $\mathrm{cond}(x, n, f)$, the term 'n' indicates the dependence of cond on the points x_j.

Lemma 5.3

$$\mathrm{cond}(x, n, f) = \frac{\sum_{j=0}^{n} |\ell_j(x) f_j|}{|p(x)|} \geq 1$$

and for any f with $|\Delta f| \leq \epsilon |f|$ we have

$$\frac{|p_f(x) - p_{f+\Delta f}(x)|}{|p_f(x)|} \leq \mathrm{cond}(x, n, f)\epsilon.$$

We introduce the relative error counter (cf. [Higham, 2002, Section 2.2])

$$\langle k \rangle := \prod_{i=1}^{k} (1+\delta_i)^{\rho_i}, \quad \rho_i = \pm 1, \quad |\delta| \le u \text{ unit roundoff.}$$

Lemma 5.4

The barycentric weights $\{\hat{\lambda}_j\}_{j=0}^n$ computed in floating point arithmetic satisfy

$$\hat{\lambda}_j = \lambda_j \langle 2n \rangle_j, \qquad j = 0, \dots, n,$$

while the computed $\hat{\ell}(x)$ satisfies $\hat{\ell}(x) = \ell(x)\langle 2n+1 \rangle$.

Theorem 5.5

The computed interpolation polynomial $\hat{p}(x)$ using the first barycentric formula satisfies

$$\hat{p}(x) = \ell(x) \sum_{j=0}^{n} \frac{\lambda_j}{x - x_j} f_j \langle 5n+5 \rangle.$$

- The statement of Theorem 5.5 is that the computed value $\hat{p}(x)$ of the interpolating polynomial at a point x is the exact value of a perturbed interpolation problem, where the perturbation is small, i.e., that interpolation via the first barycentric formula is backward stable.
- The errors are of the same form, and only $O(n)$ times larger than the errors in rounding the f_j to a floating point number.
- Applying Lemma 5.3 yields a bound for the forward error:

$$\frac{|p(x) - \hat{p}(x)|}{|p(x)|} \leq \gamma_{5n+5} \operatorname{cond}(x, n, f).$$

- If the x_j or x are not floating point numbers then there can be large relative errors in the differences $fl[fl(x_j) - fl(x_k)]$ and $fl[fl(x) - fl(x_k)]$. However, the computed $\hat{p}(x)$ can nevertheless be interpreted as the exact result corresponding to slightly perturbed x and points x_j (namely, the rounded values) and slightly perturbed points f_j; so if $p(x)$ has a large relative error, then the problem itself must be ill-conditioned with respect to variations in x and the x_j and the f_j.

Theorem 5.6

The value $\hat{p}(x)$ computed with the second barycentric formula satisfies

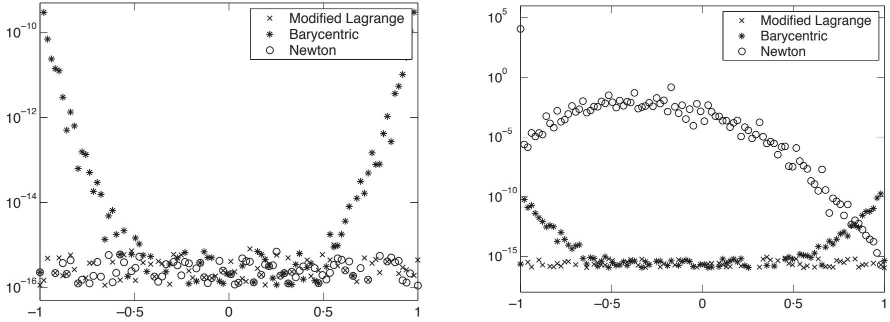
$$
\begin{aligned}
\frac{|p(x)-\hat{p}(x)|}{|p(x)|} &\le (3n+4)u\frac{\sum_{j=0}^{n}\left|\frac{\lambda_j}{x-x_j}f_j\right|}{\left|\sum_{j=0}^{n}\frac{\lambda_j}{x-x_j}f_j\right|} + (3n+2)u\frac{\sum_{j=0}^{n}\left|\frac{\lambda_j}{x-x_j}\right|}{\left|\sum_{j=0}^{n}\frac{\lambda_j}{x-x_j}\right|} + O(u^2) \\
&= (3n+4)u\,\mathbf{cond}(x,n,f) + (3n+2)u\frac{\sum_{j=0}^{n}\left|\frac{\lambda_j}{x-x_j}\right|}{\left|\sum_{j=0}^{n}\frac{\lambda_j}{x-x_j}\right|} + O(u^2) \\
&= (3n+4)u\,\mathbf{cond}(x,n,f) + (3n+2)u\,\mathbf{cond}(x,n,1) + O(u^2).
\end{aligned}
$$

- This forward error bound contains a term not present in that for the first formula. This measures the amount of cancellation in the denominator. Since the denominator is independent of f it is clear that the bound can be arbitrarily larger than $\mathbf{cond}(x, n, f)u$ for suitable choices of f and x_j.
- For example: taking $f_1 = 1$ and $f_j = 0$ for $j > 1$ gives $\mathbf{cond}(x, n, f) = 1$, while for suitable choice of the x_j the second term in the bound can become arbitrarily large.
- However, from (4.3) we see that the error bound is significantly larger than that for the first barycentric formula only if $\mathbf{cond}(x, n, 1) \gg \mathbf{cond}(x, n, f)$ a circumstance that intuitively seems unlikely.

Barycentric Interpolation Formula

Stability analysis of [Higham (2004)]

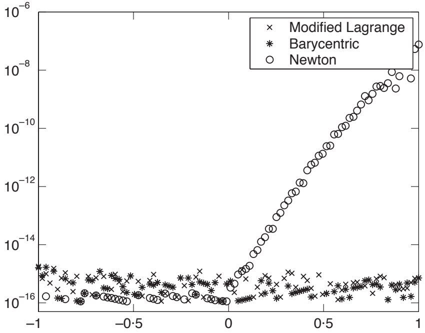
Numerical experiment: $f_j = 0$, $j = 0, \dots, n-1$, $f_n = 1$, $n = 29$, equispaced points



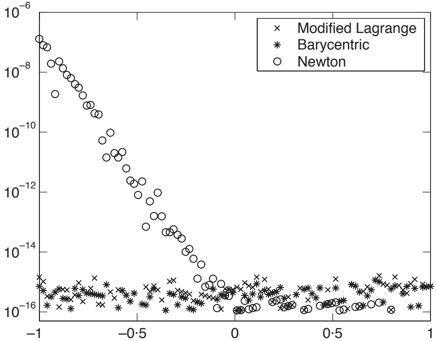
x_j in increasing order

x_j in decreasing order

Numerical experiment: Runge function, $n = 29$, Chebyshev points



x_j in increasing order



x_j in decreasing order

Contents

Theorem 6.1

For any continuous function f on $[-1,1]$ and $\epsilon > 0$ there exists a polynomial p such that $\|f - p\|_{L^\infty(-1,1)} < \epsilon$.

The original 1885 proof by Weierstraß derives the result by first extending f to a continuous function $\tilde{f}$ on $\mathbb{R}$ with compact support and then showing

$$\|\tilde{f} - p\| \le \|\tilde{f} - \tilde{f}_\epsilon\| + \|\tilde{f}_\epsilon - p\|$$

where

- $\tilde{f}_\epsilon$ is obtained from the convolution of $\tilde{f}$ with a Gaussian which is sufficiently narrow that $\|\tilde{f} - \tilde{f}_\epsilon\| < \frac{\epsilon}{2}$ and
- p is a truncated Taylor series (i.e., a polynomial) approximating the entire function $\tilde{f}_\epsilon$ suffieciently well that $\|\tilde{f}_\epsilon - p\| < \frac{\epsilon}{2}$.

Let f and g be two locally integrable[2] functions on $\mathbb{R}$. The convolution $f * g$ of f and g is a function defined on $\mathbb{R}$ by

$$(f * g)(x) = \int f(x-y)g(y)\,\mathrm{d}y = \int f(y)g(x-y)\,\mathrm{d}y = (g * f)(x)$$

provided these integrals in question exist.

[2] i.e., integrable on every compact subset of their domains of definition.

Theorem 6.2 (Generalized Young's inequality)

Let μ be a σ-finite measure on $\mathbb{R}$ as well as $1 \le p \le \infty$ and $C > 0$. Suppose K is a measurable function on $\mathbb{R}^2$ such that

$$\sup_{x\in\mathbb{R}} \int |K(x,y)| \,\mathrm{d}\mu(y) \le C \quad \text{and} \quad \sup_{y\in\mathbb{R}} \int |K(x,y)| \,\mathrm{d}\mu(x) \le C.$$

If $f \in L^p(\mathbb{R})$, then the function Tf defined by

$$(Tf)(x) := \int K(x,y) f(y) \,\mathrm{d}\mu(y)$$

is well-defined almost everywhere, lies in $L^p(\mathbb{R})$ and $\|Tf\|_{L^p(\mathbb{R})} \le C \|f\|_{L^p(\mathbb{R})}$.

As a consequence, if $f \in L^1(\mathbb{R})$ and $g \in L^p(\mathbb{R})$, $1 \le p \le \infty$, then $f * g \in L^p(\mathbb{R})$ and $\|f * g\|_{L^p(\mathbb{R})} \le \|f\|_{L^1(\mathbb{R})} \|g\|_{L^p(\mathbb{R})}$.
Proof. Apply Theorem 6.2 with $K(x,y) = f(x-y)$.

For a function f defined on $\mathbb{R}$ and $a \in \mathbb{R}$ we define $f_a(x) := f(a+x)$.

Lemma 6.3

If $1 \le p < \infty$ and $f \in L^p(\mathbb{R})$, then $\lim_{a\to 0} \|f_a - f\|_{L^p(\mathbb{R})} = 0$.

Theorem 6.4

Suppose $\phi \in L^1(\mathbb{R})$ and $\int \phi(x)\,\mathrm{d}x = c$. For each $\epsilon > 0$ define the function ϕ_ϵ by

$$\phi_\epsilon(x) := \frac{1}{\epsilon}\phi\left(\frac{x}{\epsilon}\right).$$

(a) If $f \in L^p(\mathbb{R})$, $1 \le p < \infty$, then $f * \phi_\epsilon \to cf$ in $L^p(\mathbb{R})$ as $\epsilon \to 0$.

(b) If $f \in L^\infty(\mathbb{R}($ and uniformly continuous on a set V, then $f * \phi_\epsilon \to cf$ uniformly on V as $\epsilon \to 0$.

Definition 6.5

If $f \in L^1(\mathbb{R})$, its Fourier transform $\hat{f}$ is a bounded function on $\mathbb{R}$ defined by

$$\hat{f}(\xi) = \int \mathrm{e}^{-2\pi i \xi x} f(x)\,\mathrm{d}x, \qquad \xi \in \mathbb{R}.$$

Theorem 6.6 (Convolution and Fourier transform)

If $f, g \in L^1(\mathbb{R})$, then $\widehat{(f * g)} = \hat{f} \cdot \hat{g}$.

Theorem 6.7

The Fourier transform of the function $f(x) = \mathrm{e}^{-\pi a x^2}$ with $a > 0$ is given by

$$\hat{f}(\xi) = a^{-1/2}\mathrm{e}^{-\pi \xi^2/a}.$$

Definition 6.8

The Schwartz class $\mathscr{S} = \mathscr{S}(\mathbb{R})$ is the space of all C^∞-functions on $\mathbb{R}$ which, together with all their derivatives, decay faster than any power of x as $|x| \to \infty$: $u \in \mathscr{S}$ if $u \in C^\infty(\mathbb{R})$ and

$$\sup_{x\in\mathbb{R}} |x^\alpha \partial^\beta u(x)| < \infty \qquad \text{for all } \alpha, \beta \in \mathbb{N}_0.$$

Proposition 6.9

For $f \in \mathscr{S}$ and $\beta \in \mathbb{N}_0$ there holds

(a) $\hat{f} \in C^\infty$ and $\partial^\beta \hat{f} = [(-2\pi i x)^\beta f]\hat{}$.

(b) $\widehat{\partial^\beta f} = (2\pi i \xi)^\beta \hat{f}$.

Proposition 6.10

If $f \in \mathscr{S}$, then $\hat{f} \in \mathscr{S}$.

The Cauchy problem for the heat equation in 1D with initial data f

$$\partial_t u - \partial_{xx} u = 0 \text{ on } \mathbb{R} \times (0, \infty), \qquad u(x, 0) = f(x),$$

where we assume f a function of rapid decay, can be solved by taking the Fourier transform of the heat equation with respect to x:

$$\partial_t \hat{u}(\xi, t) + 4\pi^2 \xi^2 \hat{u}(\xi, t) = 0, \qquad \hat{u}(\xi, 0) = \hat{f}(\xi).$$

This is an ordinary initial value problem which, for each value of the parameter ξ, has the solution

$$\hat{u}(\xi, t) = \hat{f}(\xi)\, \mathrm{e}^{-4\pi^2 \xi^2 t}, \qquad t > 0.$$

In other words,

$$u(x, t) = f * K_t, \quad \text{where} \quad \hat{K}_t(\xi) = \mathrm{e}^{-4\pi^2 \xi^2 t}.$$

By Theorem 6.7, this means

$$K_t(x) =: K(x,t) = (4\pi t)^{-1/2} \mathrm{e}^{-x^2/(4t)}, \qquad t > 0.$$

The function K defined on $\mathbb{R} \times (0,\infty)$ is called the Gaussian kernel or heat kernel. Note that

$$K_t(x) = t^{-1/2} K_1(t^{-1/2} x), \qquad \int K_t(x)\,\mathrm{d}x = \hat{K}_t(0) = 1.$$

By Theorem 6.4, the family $\{K_t\}_{t>0}$ is an approximation of the identity. (Set $\epsilon = \sqrt{t}$.)

Theorem 6.11

Suppose $f \in L^p(\mathbb{R})$, $1 \le p \le \infty$. Then $u(x,t) = f * K_t(x)$ satisfies $\partial_t u - \partial_{xx} u$ on $\mathbb{R} \times (0,\infty)$. If f is bounded and continuous, then u is continuous on $\mathbb{R} \times [0,\infty)$ and $u(x,0) = f(x)$. If $f \in L^p(\mathbb{R})$ where $p < \infty$, then $u(\cdot,t)$ converges to f in the L^p-norm as $t \to \infty$.

The Weierstrass Approximation Theorem

Heat kernel, remarks

- Since K_t decays rapidly as $|x| \to \infty$, $f * K_t$ exists in the interval $(0, T]$ provided only that $|f(x)| \leq C\mathrm{e}^{x^2/(4T)}$. Under this condition, one easily verifies by differentiating under the integral sign that $f * K_t$ satisfies the heat equation and approaches f uniformly on bounded sets as $t \to 0$ provided f is continuous.
- Moreover, as all derivatives of $K(x, t)$ decay rapidly as $|x| \to \infty$ we can differentiate under the integral sign to any order, and conclude that $u \in C^\infty$. Thus, the heat kernel immediately smooths out arbitrary data.

- Given $\varepsilon > 0$, by Theorem 6.11 we can find $t > 0$ such that
$$\sup_{x\in\mathbb{R}} |(\tilde{f} * K_t)(x) - \tilde{f}(x)| < \frac{\epsilon}{2}$$
- But
$$(\tilde{f} * K_t)(x) = (4\pi t)^{-1/2} \int_{\operatorname{supp} \tilde{f}} \tilde{f}(y)\, \mathrm{e}^{(x-y)^2/(4t)}\, \mathrm{d}y$$
- Since the Taylor series for e^x converges uniformly on compact sets, we can replace $\mathrm{e}^{(x-y)/(4t)}$ by a partial sum with error less than $(4\pi t)^{1/2}\epsilon/2\|\tilde{f}\|_1$ for $x \in [-1, 1]$ and $y \in \operatorname{supp} \tilde{f}$.
- Thus, $\sup_{x\in[-1,1]} |(\tilde{f} * K_t)(x) - p(x)| < \frac{\epsilon}{2}$ where
$$p(x) = (4\pi t)^{-1/2} \int_{\operatorname{supp} \tilde{f}} \tilde{f}(y) \sum_{k=0}^{K} \frac{(-1)^k}{k!} \left[\frac{(x-y)^2}{4t}\right]^k \mathrm{d}y$$
is a polynomial of degree $2K$.

∎

For a continuous function f defined on $[0,1]$ the expression

$$B_n(x) = B_{n,f}(x) := \sum_{k=0}^{n} f(\tfrac{k}{n}) \binom{n}{k} x^k (1-x)^{n-k}$$

is called the Bernstein polynomial of order n for the function f.

The Weierstrass Approximation Theorem

Bernstein's proof

For a continuous function defined on $[0, 1]$ the expression

$$B_n(x) = B_{n,f}(x) := \sum_{k=0}^{n} f\left(\tfrac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k}$$

is called the Bernstein polynomial of order n for the function f.

- The highlighted expression has the following interpretation from elementary probability theory: It represents the probability of observing exactly k successes in a binomial experiment consisting of n repeated i.i.d. Bernoulli trials with success probability x.
- In this context, the complete expression $B_n(x)$ is the expectation of a random variable whose value (payoff) is $f\left(\frac{k}{n}\right)$ when the number of successes in the above experiment is exactly $k \in \{0, 1, \dots, n\}$.
- Bernstein's proof of the Weierstrass approximation theorem establishes that this expected value, as a function of $x \in [0, 1]$, converges uniformly to f.

- Denote by X_i the random outcome of the i-th Bernoulli trial, i.e., equal to 1 with probability x and zero with probability $1-x$.
- The number of successes in n i.i.d. repeated trials is $k = X_1 + \cdots + X_n$.
- We have

$$\mathbf{E}[X_i] = x, \quad \mathbf{Var}\, X_i = \mathbf{E}\left[(X_i - \mathbf{E}[X_i])^2\right] = x(1-x),$$
$$\mathbf{E}\left[\frac{k}{n}\right] = x, \quad \mathbf{Var}\, \frac{k}{n} = \frac{x(1-x)}{n}.$$

- Intuitively, we expect the relative frequency of success k/n to approach x as n increases. By Chebyshev's inequality, for $\epsilon > 0$,

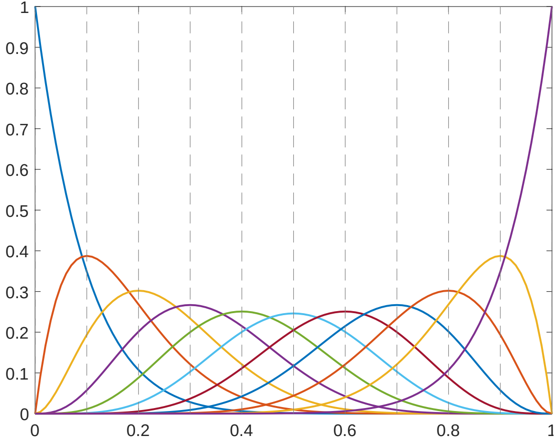
$$\mathbf{P}\left(\left|\frac{k}{n} - x\right| < \epsilon\right) \geq 1 - \frac{x(1-x)}{n\epsilon^2} \to 1 \qquad (n \to \infty).$$

- In other words

$$\sum_{\left|\frac{k}{n}-x\right|<\epsilon} \binom{n}{k} x^k (1-x)^{n-k} \to 1 \qquad (n \to \infty).$$

The Weierstrass Approximation Theorem

Bernstein basis



$$p_k(x) = \binom{n}{k} x^k (1-x)^{n-k}, \qquad k = 0, 1, \dots, n, \quad n = 10.$$

Theorem 6.12 (Bernstein, 1912)

For a function f bounded on $[0,1]$, the relation $\lim_{n\to\infty} B_n(x) = f(x)$ holds at each point of continuity x of f and holds uniformly on $[0,1]$ for $f \in C[0,1]$.

Proof:

- With $p_k(x) := \binom{n}{k} x^k (1-x)^{n-k}$, we first compute an expression for the quantity
$$T := \sum_{k=0}^{n} (k - nx)^2 p_k(x) = \sum_{k=0}^{n} \left[k(k-1) - (2nx-1)k + n^2x^2 \right] p_k(x).$$
Noting that
$$\sum_{k=0}^{n} p_k(x) = 1, \qquad \sum_{k=0}^{n} k p_k(x) = nx, \qquad \sum_{k=0}^{n} k(k-1) p_k(x) = n(n-1)x^2,$$
we conclude
$$T = n(n-1)x^2 - (2nx-1)nx + n^2x^2 = nx(1-x).$$

- Since $\left|\frac{k}{n}-x\right| \geq \delta$ implies $\frac{1}{\delta^2}\left(\frac{k}{n}-x\right)^2 \geq 1$ and since $x(1-x) \leq \frac{1}{4}$ on $[0,1]$,

$$\sum_{\left|\frac{k}{n}-x\right|\geq\delta} p_k(x) \leq \frac{1}{\delta^2} \sum_{\left|\frac{k}{n}-x\right|\geq\delta} \left(\frac{k}{n}-x\right)^2 p_k(x) \leq \frac{1}{n^2\delta^2}\sum_{k=0}^{n}(k-nx)^2 p_k(x)$$

$$= \frac{T}{n^2\delta^2} = \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}.$$

- If $|f|$ is bounded on $[0,1]$ by M and continuous at $x \in [0,1]$, then for $\epsilon > 0$ there exists $\delta > 0$ such that $|f(x) - f(x')| < \epsilon$ for all x' such that $|x - x'| < \delta$. Hence

$$|f(x) - B_n(x)| = \left|\sum_{k=0}^{n}\left[f(x) - f(\tfrac{k}{n})\right] p_k(x)\right|$$

$$= \underbrace{\sum_{\left|\frac{k}{n}-x\right|\geq\delta} \left|\left[f(x) - f(\tfrac{k}{n})\right]\right| p_k(x)}_{\leq \frac{2M}{4n\delta^2}} + \underbrace{\sum_{\left|\frac{k}{n}-x\right|<\delta} \left|\left[f(x) - f(\tfrac{k}{n})\right]\right| p_k(x)}_{\leq \epsilon \sum_{k=0}^{n} p_k(x) = \epsilon}$$

In summary, we have

$$|f(x) - B_n(x)| \le \epsilon + \frac{M}{2n\delta^2} < 2\epsilon \quad \text{for } n \text{ sufficiently large.} \tag{6.1}$$

If f is continuous on all of $[0, 1]$, then (6.1) holds for all x with δ independent of x, i.e.,

$$B_n(x) \to f(x) \quad \text{uniformly on } [0, 1].$$

■

Contents

Convergence for Differentiable functions

Theorem 7.1 (Chebyshev coefficients for differentiable functions)

For an integer $\nu \geq 0$, let f and its derivatives through $f^{(\nu-1)}$ be absolutely continuous on $[-1, 1]$ and suppose the ν-th derivative $f^{(\nu)}$ is of bounded variation V. Then for $k \geq \nu + 1$, the Chebyshev coefficients of f satisfy

$$|a_k| \leq \frac{2V}{\pi k(k-1)\cdots(k-\nu)} \leq \frac{2V}{\pi(k-\nu)^{\nu+1}}. \tag{7.1}$$

Theorem 7.2 (Convergence for differentiable functions)

If f satisfies the conditions of Theorem 7.1, with V again denoting the total variation of $f^{(\nu)}$ for som $\nu \geq 1$, then for any $n > \nu$, its Chebyshev projections satisfy

$$\|f - f_n\| \leq \frac{2V}{\pi\nu(n-\nu)^\nu} \tag{7.2}$$

and its Chebyshev interpolants satisfy

$$\|f - p_n\| \leq \frac{4V}{\pi\nu(n-\nu)^\nu} \tag{7.3}$$

Contents

Definition 10.1

We denote by p_n^* the best approximation of $f \in C[-1,1]$ by a polynomial $p \in \mathscr{P}_n$ by and set

$$E_n(f) := \inf_{p \in \mathscr{P}_n} \|f - p\|,$$

where $\|\cdot\|$ denotes the maximum norm on $[-1,1]$.

Theorem 10.2

A function f on $[-1,1]$ has a unique best approximation $p^* \in \mathscr{P}_n$. If f is real, then p^* is real, too, and in this case a polynomial $p \in \mathscr{P}_n$ is equal to p^* if and only if $f - p$ equioscillates in at least $n+2$ extreme points.

Definition 10.3 (Modulus of continuity)

Given a function $f \in C[-1,1]$, the function $\omega : [0,\infty) \to [0,\infty]$ defined by

$$\omega(\delta) = \omega(\delta; f) := \sup_{|x-y|<\delta} |f(x) - f(y)|$$

is called the modulus of continuity of f.

- Examples: for $f(x) \equiv 1$, $\omega(\delta; f) = 0$; for $f(x) = x$, $\omega(\delta; f) = \delta$; for $f(x) = x^2$, $\omega(\delta; f) = 1 - (1-\delta)^2$.
- If $0 < \delta_1 \le \delta_2$, then $\omega(\delta_1) \le \omega(\delta_2)$.
- A function f is uniformly continuous on $[-1,1]$ if and only if

$$\lim_{\delta \to 0} \omega(\delta; f) = 0.$$

Theorem 10.4 (Jackson)

For $f \in C[-1,1]$ there holds

$$E_n(f) = 6\omega\left(\frac{1}{n}\right).$$

Definition 10.5 (Total variation)

The total variation of a function $f \in C[-1,1]$ is defined by

$$V = V(f) := \sup \sum_{k=1}^{n} |f(x_k) - f(x_{k-1})|,$$

where the supremum is taken over all partitions $\{x_0 < x_1 < \cdots < x_n\}$ of $[-1,1]$. A function f is said to be of bounded variation if its total variation is finite.

- If f is differentiable and its derivative is Riemann-integrable, its total variation is given by

$$V(f) = \int_{-1}^{1} |f'(x)| \, \mathrm{d}x,$$

(i.e., the vertical component of the arc length of its graph).

Contents

- A Borel measure on $\mathbb{C}$ is any positive (nonnegative) measure defined on the Borel sets of $\mathbb{C}$.
- The support $\operatorname{supp}\mu$ of a Borel measure μ on $\mathbb{C}$ is the complement of the largest open set with measure zero.
- The Dirac measure δ_z, $z \in \mathbb{C}$, is the unit measure defined by

$$\delta_z(B) := \begin{cases} 1 & \text{if } z \in B, \\ 0 & \text{otherwise,} \end{cases} \qquad \text{for all Borel sets } B \subset \mathbb{C}.$$

- For a continuous function f on $\mathbb{C}$ there holds

$$\int f(z)\,\mathrm{d}\delta_{z_0}(z) = f(z_0).$$

Let $K \subset \mathbb{C}$ be compact in the following.

- We denote by $\mathscr{M}(K)$ the set of all finite positive Borel measures μ with $\operatorname{supp} \mu = K$ and $\mu(K) = 1$.
- The Riesz representation theorem states that for any positive[3] linear functional ℓ on $C(K)$ there exists a unique Radon measure (a Borel measure which is finite for all compact subsets of K) such that

$$\ell(f) = \int f \, \mathrm{d}\mu, \qquad \text{for all } f \in C(K).$$

- A sequence $(\mu_n)_{n\in\mathbb{N}} \subset \mathscr{M}(K)$ is said to converge to $\mu \in \mathscr{M}(K)$ in the weak*-sense if

$$\int f \, \mathrm{d}\mu_n \to \int f \, \mathrm{d}\mu \qquad \forall f \in C(K).$$

- $\mathscr{M}(K)$ is weak*-compact, i.e., every sequence $(\mu_n)_{n\in\mathbb{N}}$ has a weak*-convergent subsequence.

[3] in the sense that $\ell(f) \geq 0$ whenever $f \geq 0$

For a set of points $\{z_1, \ldots, z_n\} \subset \mathbb{C}$ we denote the associated normalized counting measure by

$$\mu_n = \frac{1}{n} \sum_{j=1}^{n} \delta_{z_j}.$$

Examples:

(1) The counting measures μ_n associated with uniformly spaced points on $K = [-1, 1]$ converge in weak* sense to $\mathrm{d}\mu(t) = \frac{1}{2}\mathrm{d}t$.

(2) The counting measures μ_n associated with the sequence of Chebyshev points on $K = [-1, 1]$ converge in weak* sense to

$$\mathrm{d}\mu(t) = \frac{\mathrm{d}t}{\pi\sqrt{1 - t^2}}.$$

For $K \subset \mathbb{C}$ compact and $\mu \in \mathscr{M}(K)$, the function $u_\mu : \mathbb{C} \to (-\infty, \infty]$ defined by

$$u_\mu(z) = \int \log \frac{1}{|z-t|} \, \mathrm{d}\mu(t) = -\int \log |z-t| \, \mathrm{d}\mu(t)$$

is called the logarithmic potential of μ.

Examples:

(1) For $\mathrm{d}\mu(t) = \frac{1}{2}\mathrm{d}t$ we obtain

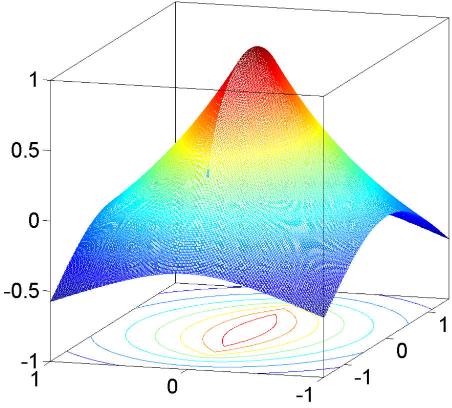
$$u_\mu(z) = 1 + \frac{1}{2} \operatorname{Re}\left[(z-1)\log(z-1) - (z+1)\log(z+1)\right], \qquad z \in \mathbb{C}.$$

(2) For $\mathrm{d}\mu(t) = \frac{\mathrm{d}}{\pi\sqrt{1-x^2}} t$ we obtain

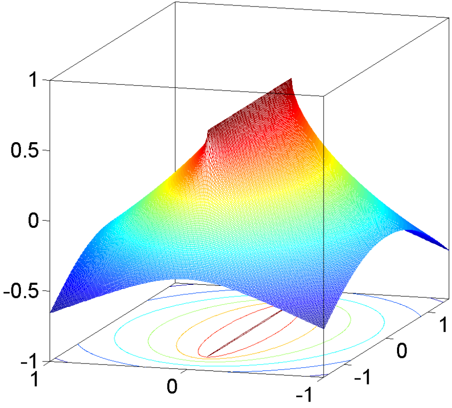
$$u_\mu(z) = \log 2 - \log \left| z + \sqrt{z^2 - 1} \right|, \qquad z \in \mathbb{C}.$$

Potential Theory

Limiting measures: equispaced and Chebyshev nodes



equispaced



Chebyshev

For $K \subset \mathbb{C}$ compact and $\mu \in \mathscr{M}(K)$, the energy of the logarithmic potential u_μ is defined as

$$I(\mu) := \int u_\mu \, \mathrm{d}\mu = \int \int \frac{1}{\log |z - t|} \, \mathrm{d}\mu(t) \, \mathrm{d}\mu(z).$$

It satisfies

$$-\infty < I(\mu) \leq \infty.$$

If the number $V(K) := \inf_{\mu \in \mathscr{M}(K)} I(\mu)$, known as the Robin constant of K, satisfies $V(K) > -\infty$, then there exists (note that $\mathscr{M}(K)$ is weak*-compact) a measure $\mu = \mu_K \in \mathscr{M}(K)$ such that $I(\mu_K) = V(K)$ known as the equilibrium distribution or equilibrium measure of K.
The number

$$\operatorname{cap}(K) := \begin{cases} \exp(-V(K)), & \text{if } V(K) > -\infty, \\ 0, & \text{otherwise} \end{cases}$$

is known as the logarithmic capacity of K.

Example: For $K = [-1, 1]$ there holds $\mu_K(t) = \frac{\mathrm{d}t}{\pi\sqrt{1-t^2}}$ and $\mathsf{cap}(K) = \frac{1}{2}$.

The following result is known as the fundamental theorem of potential theory.

Theorem 12.1 (Frostman)

For $K \subset \mathbb{C}$ compact with $\mathsf{cap}(K) > 0$ the logarithmic potential u_{μ_K} of the equilibrium measure μ_K satisfies

$$\begin{aligned} u_{\mu_K}(z) &\le V(K) \qquad \text{for all } z \in \mathbb{C} \\ u_{\mu_K}(z) &= V(K) \qquad \text{quasi-everywhere in } K. \end{aligned}$$

Here quasi-everywhere means everywhare except possibly on a set of zero capacity.

Let $K \subset \mathbb{C}$ be compact.

- For
$$\tau_n(K) := \max_{z_1,\dots,z_n \in K} \left(\prod_{1 \le j < k \le n} |z_j - z_k| \right)^{\frac{2}{n(n-1)}},$$
the transfinite diameter $\tau(K)$ of K is defined as $\tau(K) := \lim_{n\to\infty} \tau_n(K)$.
- For $E_n(K) := \min_{p \in \mathscr{P}_{n-1}} \|z^n - p_{n-1}(z)\|_{K,\infty}$, the Chebyshev constant $c(K)$ of K ist defined as
$$c(K) := \lim_{n\to\infty} E_n(K)^{1/n}.$$

Theorem 12.2 (Fekete, 1924; Szegö, 1925)

For any compact $K \subset \mathbb{C}$ there holds

$$\operatorname{cap}(K) = c(K) = \tau(K).$$

Theorem 12.3 (V. I. Krylov, 1962)

For a weak*-limiting node distribution μ and associated logarithmic potential u_μ, let f be analytic inside Γ_s, the level curve of u_μ which passes through a singularity s of f. The polynomial interpolant p_n of f associated with the n-th nodal set then converges to f uniformly inside Γ_s, diverges outside Γ_s and

$$\lim_{n\to\infty} |f(z) - p_n(z)|^{1/n} = \exp(u_\mu(s) - u_\mu(z)).$$

Contents

Equispaced Points, the Runge Phenomenon

- The Runge phenomenon refers to the 1901 paper by Carl Runge in which he presented the potential theoretic analysis of the convergence of interpolation of analytic functions. He presented an example of a meromorphic function for which interpolation at equidistant nodes on a fixed interval of analyticity diverges.
- Earlier work by Méray in 1884 and 1886 had made similar observations.
- Faber proved in 1914 that for each node sequence on a compact interval there exists a continuous function f for which the associated sequence of polynomial interpolants fail to converge to f.

Contents

Recall the Lagrange representation of the polynomial interpolant $p_n \in \mathscr{P}_n$ at distinct nodes $\{x_j\}_{j=0}^n$ with data $\{f_j\}_{j=0}^n$

$$p_n(x) = \sum_{j=0}^n f_j \, \ell_j(x)$$

in terms of the Lagrange fundamental polynomials $\{\ell_j\}_{j=0}^n$. For the interpolant $\tilde{p}_n \in \mathscr{P}_n$ on the same grid obtained from perturbed data $\{\tilde{f}_j\}_{j=0}^n$, we have

$$|p_n(x) - \tilde{p}_n(x)| \le \max_j |f_j - \tilde{f}_j| \sum_{j=0}^n |\ell_j(x)| \le \epsilon \, \lambda_n(x)$$

if $|f_j - \tilde{f}_j| \le \epsilon$ for all j, where we have introduced the Lebesgue function

$$\lambda_n(x) := \sum_{j=0}^n |\ell_j(x)|$$

associated with the interpolation nodes $\{x_0, \dots, x_n\}$.

We also define the Lebesgue constant associated with the same node set as

$$\Lambda_n := \|\lambda_n\|, \qquad \|\cdot\| = \|\cdot\|_{\infty,[-1,1]}.$$

Theorem 15.1

For a set $\{x_j\}_{j=0}^n \subset [-1,1]$ of distinct nodes, the norm of the associated polynomial interpolation operator

$$L_n : C[-1,1] \to \mathscr{P}_n, \qquad f \mapsto p_n$$

is given by $\|L_n\| = \Lambda_n$.
If p_n^* denotes the best uniform approximation of f from $\mathscr{P}_n$, then

$$\|f - p_n\| \leq (1 + \Lambda_n)\|f - p_n^*\|.$$

Theorem 15.2

The Lebesgue constants Λ_n for degree $n \geq 0$ polynomial interpolation in any set of $n+1$ distinct nodes in $[-1, 1]$ satisfy

$$\Lambda_n \geq \frac{2}{\pi} \log(n+1) + c, \tag{15.1}$$

where $c = (2/\pi)(\gamma + \log(4/\pi)) \approx 0.52125$ with Euler's constant $\gamma \approx 0.577$. For Chebyshev points they satisfy

$$\Lambda_n \leq \frac{2}{\pi} \log(n+1) + 1 \quad \text{and} \quad \Lambda_n \sim \frac{2}{\pi} \log n \quad (n \to \infty).$$

For equispaced points they satisfy

$$\Lambda_n > \frac{2^{n-2}}{n^2} \quad (n \in \mathbb{N}) \quad \text{and} \quad \Lambda_n \sim \frac{2^{n+1}}{\mathrm{e}\, n \log n} \quad (n \to \infty).$$

Theorem 15.3

The Lebesgue constants Λ_n for degree $n \geq 1$ Chebyshev projection in $[-1, 1]$ are given by

$$\Lambda_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| \frac{\sin((n + 1/2)t)}{\sin(t/2)} \right| \mathrm{d}t$$

and satisfy

$$\Lambda_n \leq \frac{4}{\pi^2} \log(n+1) + 3 \qquad \text{and} \qquad \Lambda_n \sim \frac{4}{\pi^2} \log n \quad (n \to \infty).$$

Contents

Theorem 16.1

Let $f \in C[-1,1]$ have degree n Chebyshev projection f_n, Chebyshev interpolant p_n, and best approximant p_n^*, $n \geq 1$. Then

$$\|f - f_n\| \leq \left(4 + \frac{4}{\pi^2}\log(n+1)\right)\|f - p_n^*\| \tag{16.1}$$

and

$$\|f - p_n\| \leq \left(2 + \frac{2}{\pi}\log(n+1)\right)\|f - p_n^*\|. \tag{16.2}$$

Contents

- Let $\alpha : \mathbb{R} \to \mathbb{R}$ denote a distribution function, i.e., a non-decreasing function with infinitely many points of increase which is not constant and possesses finite limits for $x \to \pm\infty$.
- Assume that all monomial moments, i.e., the Lebesgue-Stieltjes integrals
$$\mu_n := \int x^n \, \mathrm{d}\alpha(x), \qquad n \in \mathbb{N}_0,$$
are finite.
- Then a symmetric bilinear form is defined on the vector space of all (real) polynomials $\mathscr{P}$ by
$$(p, q) := \int p(x)q(x) \, \mathrm{d}\alpha(x), \qquad p, q \in \mathscr{P}. \tag{17.1}$$
- If $\mathrm{d}\alpha$ is absolutely continuous with respect to Lebesgue measure on $\mathbb{R}$, then there exists a weight function $w : \mathbb{R} \to \mathbb{R}_0^+$ such that $\mathrm{d}\alpha(x) = w(x)\mathrm{d}x$ and hence
$$(p, q) = \int p(x)q(x)w(x) \, \mathrm{d}x.$$

The classical weight functions associated with orthogonal polynomials are

supp w	$w(x)$	Name
$[-1, 1]$	1	Legendre
$[-1, 1]$	$(1 - x^2)^{-1/2}$	Chebyshev
$[-1, 1]$	$(1 - x)^\alpha (1 + x)^\beta, \ \alpha, \beta > -1$	Jacobi
$[0, \infty)$	$\exp(-x)$	Laguerre
$(-\infty, \infty)$	$\exp(-x^2)$	Hermite

The n-th moment matrix associated with the bilinear form (17.1) is given by the Hankel matrix

$$\mathbf{M}_n = \begin{bmatrix} \mu_0 & \mu_1 & \dots & \mu_{n-1} \\ \mu_1 & \mu_2 & \dots & \mu_n \\ \vdots & & & \vdots \\ \mu_{n-1} & \mu_n & \dots & \mu_{2n-2} \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad n \in \mathbb{N}.$$

Proposition 17.1

The bilinear form (17.1) is an inner product on $\mathscr{P}$ if and only if $\det \mathbf{M}_n > 0$ for all $n \in \mathbb{N}$.

Definition 17.2

A sequence of polynomials $(p_k)_{k\in\mathbb{N}_0} \subset \mathscr{P}$ is called a system of orthogonal polynomials with respect to an inner product $(\cdot,\cdot)$ on $\mathscr{P}$ if

(a) $\deg p_k = k$ for all $k \in \mathbb{N}_0$ and

(b) $(p_j, p_k) = 0$ for all $j \neq k$, $j, k \in \mathbb{N}_0$.

A system of orthogonal polynomials is unique up to normalization. Common normalizations are

- monic orthogonal polynomials characterized by a leading coefficient of one,
- orthonormal polynomials characterized by $(p_n, p_n) = 1$ for all $n \in \mathbb{N}_0$,
- polynomials taking the value one at a specific point in the support of the measure, typically the right endpoint of a bounded interval.

Lemma 17.3

The coefficient vector $\mathbf{a}^{(n)} = [a_0^{(n)}, \dots, a_{n-1}^{(n)}]^\top$ of the n-th monic orthogonal polynomial

$$p_n(x) = x^n + a_{n-1}^{(n)} x^{n-1} + \cdots + a_1^{(n)} x + a_0^{(n)}$$

with respect to an inner product $(\cdot, \cdot)$ is given by the unique solution of the linear system of equations

$$\mathbf{M}_n \mathbf{a}^{(n)} = -\mathbf{m}_n$$

with the moment matrix $\mathbf{M}_n$ and right hand side

$$\mathbf{m}_n = \begin{bmatrix} \mu_n \\ \mu_{n+1} \\ \vdots \\ \mu_{2n-1} \end{bmatrix}.$$

Theorem 17.4

A system of orthogonal polynomials satisfies a three-term recurrence relation

$$\gamma_n p_n(x) = (x - \alpha_n) p_{n-1}(x) - \beta_n p_{n-2}(x), \qquad n = 1, 2, \ldots, \tag{17.2a}$$

$$\text{with } p_{-1} := 0, \quad p_0(x) \equiv \text{ const.} \tag{17.2b}$$

The coefficients are given by

$$\alpha_n = \frac{(xp_{n-1}, p_{n-1})}{(p_{n-1}, p_{n-1})}, \qquad n = 1, 2, \ldots$$

$$\gamma_n = \frac{(xp_{n-1}, p_n)}{(p_n, p_n)}, \qquad n = 1, 2, \ldots$$

$$\beta_n = \frac{(xp_{n-2}, p_{n-1})}{(p_{n-2}, p_{n-2})} = \gamma_{n-1} \frac{(p_{n-1}, p_{n-1})}{(p_{n-2}, p_{n-2})}, \qquad n = 2, 3, \ldots, \quad \beta_1 \text{ arbitrary.}$$

Remark 17.5

(a) Rescaling a system of orthogonal polynomials $\{p_k\}_{k\ge 0}$ to $\hat{p}_k = \delta_k p_k$, $\delta_k \neq 0$, yields a system of orthogonal polynomials with associated recurrence coefficients

$$\hat{\alpha}_k = \alpha_k, \qquad \hat{\gamma}_k = \frac{\delta_{k-1}}{\delta_k}\gamma_k, \qquad k = 1, 2, \dots ,$$

$$\hat{\beta}_k = \frac{\delta_{k-1}}{\delta_{k-2}}\beta_k, \qquad k = 2, 3, \dots .$$

(b) For monic orthogonal polynomials there holds $\gamma_k = 1\,\forall k$, i.e.

$$p_{-1} = 0, \quad p_0(x) = 1, \quad p_k(x) = (x - \alpha_k)p_{k-1}(x) - \beta_k p_{k-2}(x).$$

Moreover

$$\beta_k = \frac{(p_{k-1}, p_{k-1})}{(p_{k-2}, p_{k-2})} > 0, \quad k \ge 2.$$

(c) For orthonormal polynomials there holds $\beta_k = \gamma_{k-1}$, $k \ge 2$.

We can obtain a matrix expression of the three-term recurrence (17.2a) by collecting its first n equations in a vector of polynomials

$$\mathbf{p}_n(x) := [p_0(x), p_1(x), \dots, p_{n-1}(x)]^\top,$$

resulting in the relation

$$x\mathbf{p}_n(x) = \mathbf{J}_n\mathbf{p}_n(x) + \gamma_n p_n(x)\mathbf{e}_n, \qquad \mathbf{J}_n = \begin{bmatrix} \alpha_1 & \gamma_1 & & & \\ \beta_2 & \alpha_2 & \gamma_2 & & \\ & \ddots & \ddots & \ddots & \\ & & & \gamma_{n-1} & \\ & & \beta_n & \alpha_n & \end{bmatrix}, \quad \mathbf{e}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \end{bmatrix}.$$

Rescaling the orthogonal polynomial system is reflected in the matrix relation as a diagonal scaling $\hat{\mathbf{p}}_n(x) = \mathbf{D}_n\mathbf{p}_n(x)$, $\mathbf{D}_n = \operatorname{diag}(\delta_0, \dots, \delta_{n-1})$, of the polynomial vector, resulting in a diagonal similarity transformation of the tridiagonal matrix as

$$\hat{\mathbf{J}}_n = \mathbf{D}_n\mathbf{J}_n\mathbf{D}_n^{-1}.$$

Theorem 17.6

The zeros of the orthogonal polynomials associated with the inner product (17.1) are real, simple and lie in the support interval (a, b) of $\mathrm{d}\alpha(x) = w(x)\mathrm{d}x$.

- For each zero of p_n, the matrix form of the three-term recurrence becomes an eigenvector-eigenvalue relation for the tridiagonal matrix $\mathbf{J}_n$, hence each zero of p_n is an eigenvalue of $\mathbf{J}_n$.
- The simplicity of the eigenvalues are also a consequence of the (unreduced) tridiagonal structure of $\mathbf{J}_n$.
- The diagonal scaling to obtain orthonormal polynomials results in a symmetric matrix $\mathbf{J}_n$, known as the Jacobi matrix.

Contents

Theorem 18.1

The roots of the polynomial

$$p(x) = \sum_{k=0}^{n} a_k T_x(x), \qquad a_n \neq 0,$$

are the eigenvalues of the matrix

$$C = \begin{bmatrix} 0 & 1 & & & & \\ \frac{1}{2} & 0 & \frac{1}{2} & & & \\ & \frac{1}{2} & 0 & \frac{1}{2} & & \\ & & \ddots & \ddots & \ddots & \\ & & & & & \frac{1}{2} \\ & & & & \frac{1}{2} & 0 \end{bmatrix} - \frac{1}{2a_n} \begin{bmatrix} & & & & \\ & & & & \\ & & & & \\ & & & & \\ a_0 & a_1 & a_2 & \dots & a_{n-1} \end{bmatrix}.$$

Multiple roots correspond to eigenvalues withe the same multiplicities.

Contents

The goal of numerical quadrature is to approximate integrals

$$I(f) = \int f(x)\,\mathrm{d}\alpha(x).$$

Here $\mathrm{d}\alpha(x)$ is a measure associated with distribution function α with support on a subset of the real axis as described in the chapter on orthogonal polynomials. The most common approach for constructing quadrature formulas proceeds by approximating the integrand f by a polynomial $p_n \in \mathscr{P}_n$ and then integrating the p_n exactly:

$$p_n \approx f, \quad I(f) \approx Q(f) := \int p_n(x)\,\mathrm{d}\alpha(x).$$

In terms of the Lagrange representation $p_n(x) = \sum_{j=0}^n f(x_j)\ell_j(x)$ we obtain

$$Q(f) = \sum_{j=0}^{n} w_j f(x_j) \quad \text{with quadrature weights} \quad w_j := \int \ell_j(x)\,\mathrm{d}\alpha(x). \quad (19.1)$$

By construction, interpolatory quadrature formulas with $n+1$ nodes are exact for all $p \in \mathscr{P}_n$, i.e.,

$$Q_n(p) = I(p) \quad \forall p \in \mathscr{P}_n.$$

Conversely, every $(n+1)$-point quadrature formula with exactness degree n is interpolatory.
Well-known families of interpolatory quadrature formulas are the

- Newton-Cotes formulas, characterized by equispaced nodes. These include the midpoint rule ($n = 0$), the trapezoidal rule ($n = 1$), Simpson's rule ($n = 2$), the 3/8-rule ($n = 3$), Milne's rule ($n = 4$) and Weddle's rule ($n = 6$). The degree of exactness of is actually $n+1$ for n even[4]
  For $n \geq 7$ these rules have negative weights (which also grow exponentially with n), leading to numerical instability (cancellation) and, as shown in [Polya, 1933], a non-convergent sequence of rules even for analytic functions.
- The Clenshaw-Curtis formulas result from choosing Chebyshev nodes.

[4] i.e., for an *odd* number of nodes

For a given node set $\{x_j\}_{j=0}^n$, interpolatory quadrature formulas determine the *weights* to achieve a degree of exactness n. Gauss quadrature rules additionally choose the *nodes* in a clever way to achieve a higher degree of exactness.

Theorem 19.1 (Jacobi, 1826)

The quadrature rule (19.1) possesses degree of exactness $d = n + m$ for $m \in \mathbb{N}_0$ if and only if

(a) (19.1) is interpolatory and

(b) the nodal polynomial $\omega_{n+1}(x) = \prod_{j=0}^n (x - x_j)$ is orthogonal to $\mathscr{P}_{m-1}$ with respect to the inner product

$$(p, q) = \int p(x)q(x)\,\mathrm{d}\alpha(x), \quad p, q \in \mathscr{P}. \tag{19.2}$$

Remark 19.2

Maximal achievable exactness degree is $d = 2n + 1$ corresponding to $m = n + 1$.

- From Theorem 19.1 we immediately conclude that an optimal choice of quadrature (interpolation) nodes results when the associated nodal polynomial ω_{n+1} is orthogonal to $\mathscr{P}_n$.
- With the zeros of the Legendre polynomial P_{n+1} as nodes we obtain

$$Q_n(p) = \int_{-1}^{1} p(x)\frac{1}{2}\,\mathrm{d}x \quad \forall p \in \mathscr{P}_{2n+1}. \qquad \text{(Gauss-Legendre quadrature)}$$

  Similarly, with the zeros of the Jacobi polynomials $P_{n+1}^{(\alpha,\beta)}$,

$$Q_n(p) = \int_{-1}^{1} p(x)(1-x)^{\alpha}(1+x)^{\beta}\,\mathrm{d}x \quad \forall p \in \mathscr{P}_{2n+1}.$$

(Gauss-Jacobi quadrature)

- In the same way, Gauss-Laguerre and Gauss-Hermite quadrature formulas are obtained for the intervals $(0,\infty)$ and $(-\infty,\infty)$, respectively.

Theorem 19.3

For the recurrence coefficients $\alpha_k, \beta_k, k \geq 1$ of the monic orthogonal polynomials with respect to (19.2), define the sequence of Jacobi-matrices

$$J_n = \begin{bmatrix} \alpha_1 & \sqrt{\beta_2} & & \\ \sqrt{\beta_2} & \alpha_2 & \ddots & \\ & \ddots & \ddots & \sqrt{\beta_n} \\ & & \sqrt{\beta_n} & \alpha_n \end{bmatrix} \in \mathbb{R}^{n\times n}, \quad (n \in \mathbb{N}), \textit{ then}$$

(a) the nodes of the Gauss quadrature rule of order $n-1$ associated with (19.2) are the n (distinct) eigenvalues of J_n and

(b) if $\mathbf{u}_j$ denote normalized eigenvectors of J_n associated with eigenvalues λ_j, i.e. $J_n\mathbf{u}_j = \lambda_j\mathbf{u}_j$, $\|\mathbf{u}_j\|_2 = 1\,(j = 1, \dots, n)$ then the associated weights w_j are given by

$$w_j = \beta_0[\mathbf{u}_j]_1^2 \quad (j = 1, \dots, n), \qquad\qquad \beta_0 := \int \mathrm{d}\alpha(x).$$

- This observation leads to an elegant algorithm proposed by [Golub & Welsch, 1969] for generating the nodes and weights of classical Gauss quadrature rules as well as those for any measure $\mathrm{d}\alpha$ for which one can generate the recurrence coefficients.
- Simple modifications of Jacobi measures lead to quadrature rules with one or both endpoints ar nodes known as Gauss-Radau or Gauss-Lobatto rules, respectively, which can also be constructed this way by a low-rank modification of the Jacobi matrix.
- Since it involves computing the eigenvalues of a symmetric tridiagonal $n \times n$ matrix, this algorithm has complexity $O(n^2)$.
- More recently, [Glaser, Liu & Rokhlin, 2007] introduced an algorithm with superior complexity $O(n)$, which was subsequently improved in [Hale & Townsend, 2012]. This allows the stable and fast computation of Gauss quadrature rules of essentially any order.

Proposition 19.4

The Gauss-Chebyshev quadrature nodes and weights for weight function $w(x) = (1-x^2)^{-1/2}$ are given by

$$x_j = \cos\frac{(2j+1)\pi}{2(n+1)}, \quad w_j = \frac{\pi}{n}, \qquad j = 0, 1, \dots, n.$$

The integral over $[-1, 1]$ of a Chebyshev polynomial of odd degree is zero, and for even degree it is

$$\int_{-1}^{1} T_k(x)\,\mathrm{d}x = \frac{2}{1-k^2}. \tag{19.3}$$

Proposition 19.5

The integral of a polynomial $p_n \in \mathscr{P}_n$ in Chebyshev representation is given by

$$\int_{-1}^{1} p_n(x)\,\mathrm{d}x = \sum_{\substack{k=0 \\ k \text{ even}}}^{n} \frac{2a_k}{1-k^2}, \qquad p_n(x) = \sum_{k=0}^{n} a_k T_k(x).$$

Theorem 19.6 (Quadrature for analytic integrands)

Let a function f be analytic in $[-1, 1]$ and analytically continuable to the open Bernstein ellipse E_ρ ($\rho > 1$) where it satisfies $|f(z)| \leq M$ for some M. Then $(n+1)$-point Clenshaw–Curtis quadrature with $n \geq 2$ applied to f satisfies

$$|I(f) - Q_n(f)| \leq \frac{64}{15} \frac{M\rho^{1-n}}{\rho^2 - 1} \tag{19.4}$$

and $(n+1)$-point Gauss quadrature with $n \geq 1$ satisfies

$$|I(f) - Q_n(f)| \leq \frac{64}{15} \frac{M\rho^{-2n}}{\rho^2 - 1}. \tag{19.5}$$

The factor ρ^{1-n} in (19.4) can be improved to ρ^{-n} if n is even, and the factor $64/15$ can be improved to $144/35$ if $n \geq 4$ in (19.4) or $n \geq 2$ in (19.5).

Theorem 19.7 (Quadrature for differentiable integrands)

(a) For any $f \in C[-1,1]$, both Clenshaw–Curtis and Gauss quadratures $Q_n(f)$ converge to the integral $I(f)$ as $n \to \infty$.

(b) For an integer $\nu \geq 1$, let f and its derivatives through $f^{(\nu-1)}$ be absolutely continuous on $[-1,1]$ and suppose the ν-th derivative $f^{(\nu)}$ is of bounded variation V. Then $(n+1)$-point Clenshaw–Curtis quadrature applied to f satisfies

$$|I(f) - Q_n(f)| \leq \frac{32}{15} \frac{V}{\pi\nu(n-\nu)^{\nu}} \quad \text{for } n > \nu \tag{19.6}$$

and $(n+1)$-point Gauss quadrature satisfies

$$|I(f) - Q_n(f)| \leq \frac{32}{15} \frac{V}{\pi\nu(n-2\nu-1)^{2\nu+1}} \quad \text{for } n > 2\nu + 1. \tag{19.7}$$

Theorem 19.8

Under the hypotheses of Theorem 19.7, the same conclusion (19.7) also holds for $(n+1)$-point Clenshaw–Curtis quadrature:

$$|I(f) - Q_n(f)| \le \frac{32}{15} \frac{V}{\pi\nu(n-2\nu-1)^{2\nu+1}}. \tag{19.8}$$

The only difference is that this bound applies for all sufficiently large n (depending on ν but not f) rather than for $n > 2\nu + 1$.

Proposition 19.9

The barycentric weights λ_j for polynomial interpolation at Legendre points can be written as

$$\lambda_j = (-1)^j \sqrt{(1 - x_j^2) w_j},$$

where $\{x_j\}$ and $\{w_j\}$ are the nodes and weights for $(n+1)$-point Gauss-Legendre quadrature.

Gauss and Clenshaw-Curtis Quadrature

ATAP Exercise 19.8

- Approximating $I(f) = \int_0^1 f(x)\,\mathrm{d}x$ by $Q_n(f) = \int_0^1 B_{n,f}(x)\,\mathrm{d}x$, where $B_{n,f}$ denotes the Bernstein polynomial of degree n associated with f, results in the equal-weight quadrature formula

$$Q_n(f) = \frac{1}{n+1} \sum_{j=0}^{n} f\left(\tfrac{j}{n}\right),$$

  as can be verified by a simple induction.
- The degree of exactness is 1 for $n \geq 1$, hence Q_n is not interpolatory.
- For integrands of bounded variation, it follows from Koksma's inequality that the error of this quadrature formula is $O(n^{-1})$.

Theorem 19.10 (Koksma's inequality)

Given a function of bounded variation $V(f)$ on $[0,1]$ and a point set $\{x_j\}_{j=1}^n \subset [0,1]$ with star discrepancy D_n^*, then

$$\left| \frac{1}{n} \sum_{j=1}^{n} f(x_j) - \int_0^1 f(x)\,\mathrm{d}x \right| \leq V(f)\, D_n^*.$$

The star-discrepancy of a point set $\{x_j\}_{j=1}^n \subset [0,1]$ is defined at

$$D_n^*(x_1,\dots,x_n) := \sup_{\alpha\in(0,1]} \left| \frac{A((0,\alpha];\{x_j\})}{n} - \alpha \right|.$$

Here $A((0,\alpha])$ denotes the number of points of the set $\{x_j\}_{j=1}^n$ contained in $(0,\alpha]$.

The following result of Niederreiter allows us to calculate the star-discrepancy of an equispaced point set:

Theorem 19.11 (Niederreiter)

Let $x_1, \leq x_2 \leq \cdots \leq x_n$ be n numbers in $[0,1]$. Then their star-discrepancy D_n^* is given by

$$D_n^* = \max_{j=1,\dots,n} \max \left\{ \left| x_j - \frac{j}{n} \right|, \left| x_j - \frac{j-1}{n} \right| \right\} = \frac{1}{2n} + \max_{j=1,\dots,n} \left| x_j - \frac{2j-1}{2n} \right|.$$

Contents

- To approximate a real-valued function f on $[-1,1]$ by a polynomial of degree $n \geq 0$, suppose f has an absolutely convergent Chebyshev expansion

$$f(x) = \sum_{k=0}^{\infty} a_k T_k(x).$$

- For now, suppose a_{n+1} is the first nonzero coefficient and that the expansion is finite, terminating at $k = N \geq n+1$:

$$f(x) = \sum_{k=n+1}^{N} a_k T_k(x).$$

- Now make the familiar substitution $x = \frac{1}{2}(z + z^{-1}) = \operatorname{Re} z$, $|z| = 1$, and define F on $|z| = 1$ by $F(z) = F(z^{-1}) = f(x)$, leading to a Laurent expansion of F as

$$F(z) = \frac{1}{2} \sum_{k=n+1}^{N} a_k (z^k + z^{-k}).$$

Carathéodory-Fejér Approximation

Setting

- Separating $F(z) = G(z) + G(z^{-1})$ into its analytic part

$$G(z) = \frac{1}{2} \sum_{k=n+1}^{N} a_k z^k$$

  and co-analytic part $G(z^{-1})$, we note that the former can be analytically continued to $|z| \leq 1$ and the latter to $|z| \geq 1$.
- Consider the problem of approximating G on $|z| = 1$ by a function defined by a series

$$\tilde{P}(z) = \frac{1}{2} \sum_{k=-\infty}^{n} b_k z^k$$

  converging in $|z| \geq 1$.

Theorem 20.1 (Carathéorory & Fejér (1911); Schur (1918))

The approximation problem described on the previous slide has a unique solution $\tilde{P}$ given by the error formula

$$(G - \tilde{P})(z) = \lambda z^{n+1} \frac{u(z)}{\overline{u(z)}}, \tag{20.1}$$

where λ is the eigenvalue of largest magnitude of the Hankel matrix

$$H = \begin{bmatrix} a_{n+1} & a_{n+2} & a_{n+3} & \dots & a_N \\ a_{n+2} & a_{n+3} & & & \\ a_{n+3} & & \iddots & & \\ \vdots & & & & \\ a_N & & & & \end{bmatrix} \text{ with associated real eigenvector } \begin{bmatrix} u_0 \\ u_1 \\ \vdots \\ u_{N-n-1} \end{bmatrix}$$

and $u(z) = u_0 + u_1 z + \cdots + u_{N-n-1} z^{N-n-1}$. The function $G - \tilde{P}$ maps the unit circle to a circle of radius $|\lambda|$ and winding number $\geq n + 1$ with equality holding if $|\lambda| > |\mu|$ for all other $\mu \in \Lambda(H)$.

- To construct a *polynomial* approximant from $\tilde{P}$, note that, since $G - \tilde{P}$ maps $|z| = 1$ to a circle of winding number $\geq n + 1$, its real part (times 2)

$$(G - \tilde{P})(z) + (G - \tilde{P})(z^{-1})$$

maps $[-1, 1]$ to an error curve which equioscillates $\geq n + 2$ times.
- This suggests $\tilde{p}(x) := \tilde{P}(z) + \tilde{P}(z^{-1})$ as an approximation with the correct error equioscillation behavior. However, $\tilde{p}$ is not a polynomial of degree n.
- By truncating the Laurent expansion of $\tilde{P}$ to $P_{\mathrm{CF}}(z) := \frac{1}{2}\sum_{k=-n}^{n} b_k z^k$ with real part

$$p_{\mathrm{CF}}(x) := P_{\mathrm{CF}}(z) + P_{\mathrm{CF}}(z^{-1}) = \frac{1}{2}\sum_{k=-n}^{n}(b_k + b_{-k})z^k,$$

we obtain a polynomial approximation $p_{\mathrm{CF}} \in \mathscr{P}_n$ whose error curve $f - p_{\mathrm{CF}}$ will nearly match the equioscillation behavior of $f - \tilde{p}$ on $[-1, 1]$ if the truncated terms are small.

- To understand why this approximation can be expected to be good, suppose f is analytic on $[-1, 1]$ with geometrically decaying Chebyshev coefficients $a_k = O(\rho^k)$.
- Then the dominant degree $n + 1$ term of f is of order ρ^{-n-1} and the terms $b_n, b_{n-1}, \dots, b_{-n}$ are of orders $\rho^{-n-2}, \rho^{-n-3}, \dots, \rho^{-3n-2}$, which suggests an error of order ρ^{-3n-3} is committed by the truncation from $\tilde{p}$ to p_{CF}.
- This is generally small compared to, e.g., the error of best approximation $\|f - p^*\|$, which is of order ρ^{-n-1}.

Theorem 20.2 (Gutknecht & Trefethen (1982))

For any fixed $m \geq 0$, let f have a Lipschitz continuous derivative of order 3m+3 on $[-1, 1]$ with a nonzero $(m+1)$st derivative at $x = 0$, and for each $s \in (0, 1]$, let p^* and p_{CF} be the best and CF approximations of degree m to $f(sx)$ in $[-1, 1]$, respectively. Then as $s \to 0$,

$$\|f - p^*\| = O(s^{m+1}), \tag{20.2}$$

$$\|f - p^*\| \neq O(s^{m+2}), \tag{20.3}$$

$$\|p_{\text{CF}} - p^*\| = O(s^{3m+4}). \tag{20.4}$$

- Theorem 20.1 still applies if f is not a polynomial of degree N but has an absolutely convergent Chebyshev series. In this case H is the matrix representation of a compact operator on ℓ^2 or ℓ^1 and $u(z)$ is defined by an infinite series of eigenvector entries.
  [Hayashi, Trefethen & Gutknecht, 1990].
- The theory of CF approximation also extends to *rational* in place of polynomial approximation. Seminal work here is attributed to the Ukranian mathematicians [Adamayan, Arov & Krein, 1971].

Contents

Spectral Methods

- Spectral methods refers to a class of methods for approximating the solution of differential equations.
- Like finite element methods, they construct solution approximations in finite dimensional function spaces. Approximations are selected by applying sufficiently many constraints, either by imposing a variational equality or requiring the equation to hold *exactly* at a finite number of collocation points.
- Unlike finite element methods, based on piecewise polynomials as trial functions, spectral methods use global algebraic or trigonometric polynomials.
- Spectral methods converge exponentially when the solutions are analytic, and approximate derivatives of the solution to the same order.
- They are difficult to apply to non-separable geometries and are therefore commonly used in turbulence simulation and atmospheric simlulations, where domains are simple and the solutions smooth.
- Fundamental techniques for solving random differential equations as arise in uncertainty quantification, sometimes known as polynomial chaos expansions going back to Norbert Wiener in the 1930s, are spectral methods based originally on Hermite expansions.

Theorem 21.1

Let a function f be analytic in $[-1,1]$ and analytically continuable to the closed Bernstein ellipse $\overline{E_\rho}$ for some $\rho > 1$. Then for each integer $\nu \geq 0$, the νth derivatives of the Chebyshev projections f_n and interpolants p_n satisfy as $n \to \infty$

$$\|f^{(\nu)} - f_n^{(\nu)}\| = O(\rho^{-n}), \quad \|f^{(\nu)} - p_n^{(\nu)}\| = O(\rho^{-n}) \tag{21.1}$$

cf. [Tadmor, 1986].

- In spectral collocation methods, the approximate solution of a differential equation (DE) is sought in a finite-dimensional space of trial functions and determined uniquely by requiring the approximation to solve the DE exactly at an appropriate number of collocation points.
- If the approximate solution is represented as the interpolant

$$u_n(x) = \sum_{j=1}^{n} \alpha_j \, \ell_j(x)$$

of its function values $u_n(x_j) = \alpha_j$ at the collocation points $\{x_j\}_{j=1}^n$, then applying the differential operator $\mathrm{d}/\mathrm{d}x$ to u_n and evaluating at a collocation point x_k yields

$$\frac{\mathrm{d}}{\mathrm{d}x} u_n(x_k) = \sum_{j=1}^{n} \alpha_j \, \ell_j'(x_k).$$

- Thus, the linear mapping that takes the vector of function values $u_n(x_k)$ to the derivatives $u_n'(x_k)$ is represented by the differentiation matrix $D = [\ell_j'(x_k)]_{k,j=1}^n$.

- Closed form representation of differentiation matrices can be derived. E.g. for the first derivative we have

$$\ell_j'(x_k) = \begin{cases} \frac{\lambda_j}{\lambda_k(x_k - x_j)} & j \neq k, \\ \frac{x_j}{1-x_j^2} & j = k, \end{cases}$$

  where the λ_j denote the barycentric weights associated with the collocation points.