Lecture 11: Interpolation Error Bounds

2.5. Convergence of interpolants.

Interpolation can be used to generate low-degree polynomials that approximate a complicated function over the interval $[a, b]$. One might assume that the more data points that are interpolated (for a fixed $[a, b]$), the more accurate the resulting approximation. In this lecture, we address the behavior of the maximum error

$$\max_{x\in[a,b]} |f(x) - p_n(x)|$$

as the number of interpolation points—hence, the degree of the interpolating polynomial—is increased. We begin with a theoretical result.

Theorem (Weierstrass Approximation Theorem). Suppose $f \in C[a, b]$. For any $\varepsilon > 0$ there exists some polynomial p_n of finite degree n such that $\max_{x\in[a,b]} |f(x) - p_n(x)| \le \varepsilon$.

Unfortunately, we do not have time to prove this in class.† As stated, this theorem gives no hint about what the approximating polynomial looks like, whether p_n interpolates f at $n + 1$ points, or merely approximates f well throughout $[a, b]$, nor does the Weierstrass theorem describe the accuracy of a polynomial for a specific value of n (though one could gain insight into such questions by studying the constructive proof).

On the other hand, for the interpolation problem studied in the preceding lectures, we can obtain a specific error formula that gives a bound on $\max_{x\in[a,b]} |f(x) - p_n(x)|$. From this bound, we can deduce if interpolating f at increasingly many points will eventually yield a polynomial approximation to f that is accurate to any specified precision.

Theorem (Interpolation Error Bound). Suppose $f \in C^{n+1}[a, b]$ and let $p_n \in \mathcal{P}_n$ denote the polynomial that interpolates $\{(x_j, f(x_j)\}_{j=0}^n$ with $x_j \in [a, b]$ for $j = 0, \ldots, n$. The for every $x \in [a, b]$ there exists $\xi \in [a, b]$ such that

$$f(x) - p_n(x) = \frac{f^{(n+1)}(\xi)}{(n+1)!} \prod_{j=0}^{n} (x - x_j).$$

This result yields a bound for the worst error over the interval $[a, b]$:

$$\max_{x\in[a,b]} |f(x) - p_n(x)| \le \left(\max_{\xi\in[a,b]} \frac{|f^{(n+1)}(\xi)|}{(n+1)!} \right) \left(\max_{x\in[a,b]} \prod_{j=0}^{n} |x - x_j| \right). \tag{11.1}$$

We shall carefully prove this essential result; it will repay the effort, for this theorem becomes the foundation upon which we shall build the convergence theory for piecewise polynomial approximation and interpolatory quadrature rules for definite integrals.

Proof. Consider some arbitrary point $\widehat{x} \in [a, b]$. We seek a descriptive expression for the error $f(\widehat{x}) - p_n(\widehat{x})$. If $\widehat{x} = x_j$ for some $j \in \{0, \ldots, n\}$, then $f(\widehat{x}) - p_n(\widehat{x}) = 0$ and there is nothing to prove. Thus, suppose for the rest of the proof that $\widehat{x}$ is not one of the interpolation points.

†The typical proof is a construction based on Bernstein polynomials; see, e.g., Kincaid and Cheney, *Numerical Analysis*, 3rd edition, pages 320–323. This result can be generalized to the Stone–Weierstrass Theorem, itself a special case of Bishop's Theorem for approximation problems in operator algebras; see e.g., §5.6–§5.8 of Rudin, *Functional Analysis*, second ed., McGraw Hill, 1991.

To describe $f(\widehat{x}) - p_n(\widehat{x})$, we shall build the polynomial of degree $n+1$ that interpolates f at $x_0, \ldots, x_n$, and also $\widehat{x}$. Of course, this polynomial will give zero error at $\widehat{x}$, since it interpolates f there. From this polynomial we can extract a formula for $f(\widehat{x}) - p_n(\widehat{x})$ by measuring how much the degree $n+1$ interpolant improves upon the degree-n interpolant p_n at $\widehat{x}$.

Since we wish to understand the relationship of the degree $n+1$ interpolant to p_n, we shall write that degree $n+1$ interpolant in a manner that explicitly incorporates p_n. Given this setting, use of the Newton form of the interpolant is natural; i.e., we write the degree $n+1$ polynomial as

$$p_n(x) + \lambda \prod_{j=0}^{n} (x - x_j)$$

for some constant λ chosen to make the interpolant exact at $\widehat{x}$. For convenience, we write

$$w(x) \equiv \prod_{j=0}^{n} (x - x_j)$$

and then denote the error of this degree $n+1$ interpolant by

$$\phi(x) \equiv f(x) - (p_n(x) + \lambda w(x)).$$

To make the polynomial $p_n(x) + \lambda w(x)$ interpolate f at $\widehat{x}$, we shall pick λ such that $\phi(\widehat{x}) = 0$. The fact that $\widehat{x} \not\in \{x_j\}_{j=0}^{n}$ ensures that $w(\widehat{x}) \neq 0$, and so we can force $\phi(\widehat{x}) = 0$ by setting

$$\lambda = \frac{f(\widehat{x}) - p_n(\widehat{x})}{w(\widehat{x})}.$$

Furthermore, since $f(x_j) = p_n(x_j)$ and $w(x_j) = 0$ at all the $n+1$ interpolation points $x_0, \ldots, x_n$, we also have $\phi(x_j) = f(x_j) - p_n(x_j) - \lambda w(x_j) = 0$. Thus, ϕ is a function with at least $n+2$ zeros in the interval $[a, b]$. Rolle's Theorem[‡] tells us that between every two consecutive zeros of ϕ, there is some zero of ϕ'. Since ϕ has at least $n+2$ zeros in $[a, b]$, ϕ' has at least $n+1$ zeros in this same interval. We can repeat this argument with ϕ' to see that ϕ'' must have at least n zeros in $[a, b]$. Continuing in this manner with higher derivatives, we eventually conclude that $\phi^{(n+1)}$ must have at least one zero in $[a, b]$; we denote this zero as ξ, so that $\phi^{(n+1)}(\xi) = 0$.

We now want a more concrete expression for $\phi^{(n+1)}$. Note that

$$\phi^{(n+1)}(x) = f^{(n+1)}(x) - p_n^{(n+1)}(x) - \lambda w^{(n+1)}(x).$$

Since p_n is a polynomial of degree n or less, $p_n^{(n+1)} \equiv 0$. Now observe that w is a polynomial of degree $n+1$. We could write out all the coefficients of this polynomial explicitly, but that is a bit tedious, and we do not need all of them. Simply observe that we can write $w(x) = x^{n+1} + q(x)$, for some $q \in \mathcal{P}_n$, and this polynomial q will vanish when we take $n+1$ derivatives:

$$w^{(n+1)}(x) = \left(\frac{d^{n+1}}{dx^{n+1}}\, x^{n+1} \right) + q^{(n+1)}(x) = (n+1)! + 0.$$

[‡]Recall the Mean Value Theorem from calculus: Given $d > c$, suppose $f \in C[c, d]$ is differentiable on (c, d). Then there exists some $\eta \in (c, d)$ such that $(f(d) - f(c))/(d - c) = f'(\eta)$. Rolle's Theorem is a special case: If $f(d) = f(c)$, then there is some point $\eta \in (c, d)$ such that $f'(\eta) = 0$.

Assembling the pieces, $\phi^{(n+1)}(x) = f^{(n+1)}(x) - \lambda\,(n+1)!$. Since $\phi^{(n+1)}(\xi) = 0$, we conclude that

$$\lambda = \frac{f^{(n+1)}(\xi)}{(n+1)!}.$$

Substituting this expression into $0 = \phi(\widehat{x}) = f(\widehat{x}) - p_n(\widehat{x}) - \lambda w(\widehat{x})$, we obtain

$$f(\widehat{x}) - p_n(\widehat{x}) = \frac{f^{(n+1)}(\xi)}{(n+1)!}\prod_{j=0}^{n}(\widehat{x} - x_j). \qquad \blacksquare$$

This error bound has strong parallels to the remainder term in Taylor's formula. Recall that for sufficiently smooth h, the Taylor expansion of f about the point x_0 is given by

$$f(x) = f(x_0) + (x - x_0)f'(x_0) + \cdots + \frac{(x - x_0)^k}{k!}f^{(k)}(x_0) + \text{REMAINDER}.$$

Ignoring the remainder term at the end, note that the Taylor expansion gives a polynomial model of f, but one based on local information about f and its derivatives, as opposed to the polynomial interpolant, which is based on global information, but only about f, not its derivatives.

An interesting feature of the interpolation bound is the polynomial $w(x) = \prod_{j=0}^{n}(x - x_j)$. This quantity plays an essential role in approximation theory, and also a closely allied subdiscipline of complex analysis called *potential theory.* Naturally, one might wonder what choice of points $\{x_j\}$ minimizes $|w(x)|$: We will revisit this question when we study approximation theory in the near future. For now, we simply note that the points that minimize $|w(x)|$ over $[a, b]$ are called *Chebyshev points*, which are clustered more densely at the ends of the interval $[a, b]$.

Example ($f(x) = \sin(x)$). We shall apply the interpolation bound to $f(x) = \sin(x)$ on $x \in [-5, 5]$. Since $f^{(n+1)}(x) = \pm\sin(x)$ or $\pm\cos(x)$, we have $\max_{x\in[-5,5]} |f^{(n+1)}(x)| = 1$ for all n. The interpolation result we just proved then implies that *for any choice of distinct interpolation points in* $[-5, 5]$,

$$\prod_{j=0}^{n} |x - x_j| < 10^{n+1},$$

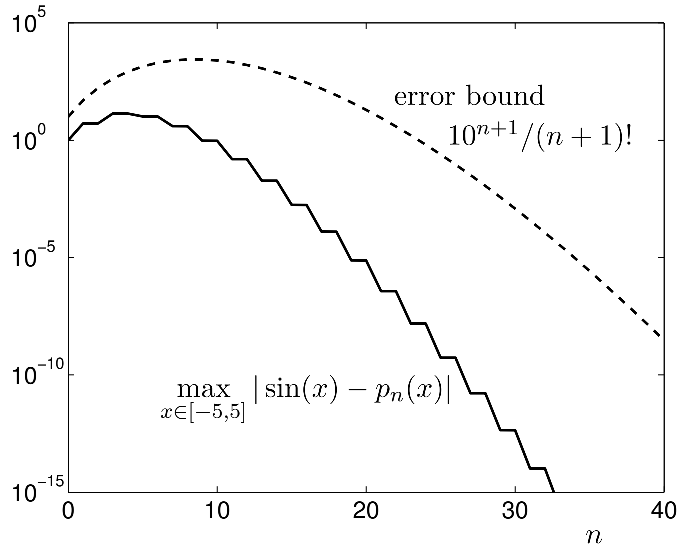
the worst case coming if all the interpolation points are clustered at an end of the interval $[-5, 5]$. Now our theorem ensures that

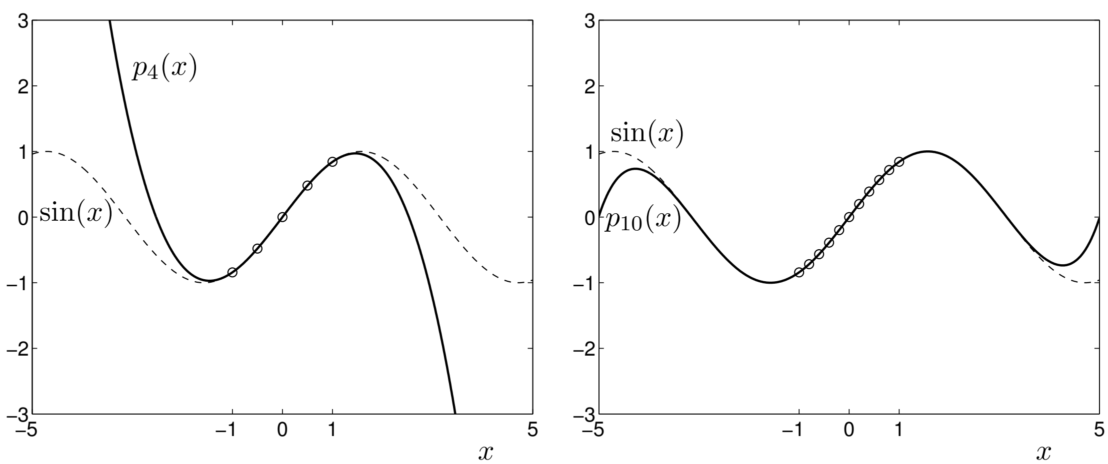
$$\max_{x\in[-5,5]} |\sin(x) - p_n(x)| \le \frac{10^{n+1}}{(n+1)!}.$$

For small values of n, this bound will be very large, but eventually $(n+1)!$ grows much faster than 10^{n+1}, so we conclude that our error must go to zero as $n \to \infty$ *regardless of where in* $[-5, 5]$ *we place our interpolation points*! The error bound is shown in the first plot below.

Consider the following specific example: Interpolate $\sin(x)$ at points uniformly selected in $[-1, 1]$. At first glance, you might think there is no reason that we should expect our interpolants p_n to converge to $\sin(x)$ for all $x \in [-5, 5]$, since we are only using data from the subinterval $[-1, 1]$, which is only 20% of the total interval and does not even include one entire period of the sine function. (In fact, $\sin(x)$ attains neither its maximum nor minimum on $[-1, 1]$.) Yet the error bound we proved above ensures that the polynomial interpolant must converge throughout $[-5, 5]$.

This is illustrated in the first plot below. The next plots show the interpolants $p_4(x)$ and $p_{10}(x)$ generated from these interpolation points. Not surprisingly, these interpolants are most accurate near $[-1, 1]$, the location of the interpolation points (shown as circles), but we still see convergence well beyond $[-1, 1]$, in the same way that the Taylor expansion for $\sin(x)$ at $x = 0$ will converge everywhere.





Example (Runge's Example). The error bound (11.1) suggests those functions for which interpolants might fail to converge as $n \to \infty$: beware if higher derivatives of f are large in magnitude over the interpolation interval. The most famous example of such behavior is due to Carl Runge, who studied convergence of interpolants for $f(x) = 1/(1+x^2)$ on the interval $[-5, 5]$. This function looks beautiful —it resembles a bell curve, with no singularities in sight—but successive derivatives expose its flaw:

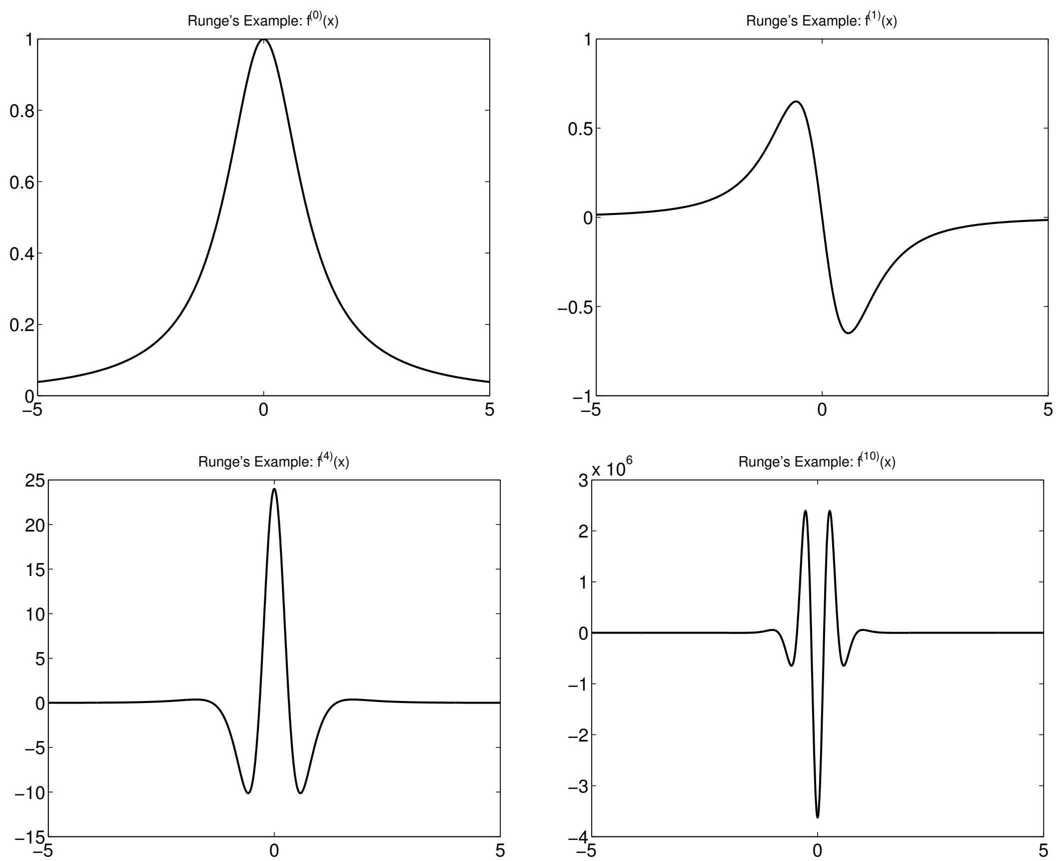
$$f'(x) = -\frac{2x}{(1+x^2)^2}, \qquad f''(x) = \frac{8x^2}{(1+x^2)^3} - \frac{2}{(1+x^2)^2}, \qquad f'''(x) = -\frac{48x^3}{(1+x^2)^4} + \frac{24x}{(1+x^2)^3}.$$

$$f^{(iv)}(x) = \frac{348x^4}{(1+x^2)^5} - \frac{288x^2}{(1+x^2)^4} + \frac{24}{(1+x^2)^3}, \qquad f^{(vi)}(x) = \frac{46080x^6}{(1+x^2)^7} - \frac{57600x^4}{(1+x^2)^6} + \frac{17280x^2}{(1+x^2)^5} - \frac{720}{(1+x^2)^4}.$$

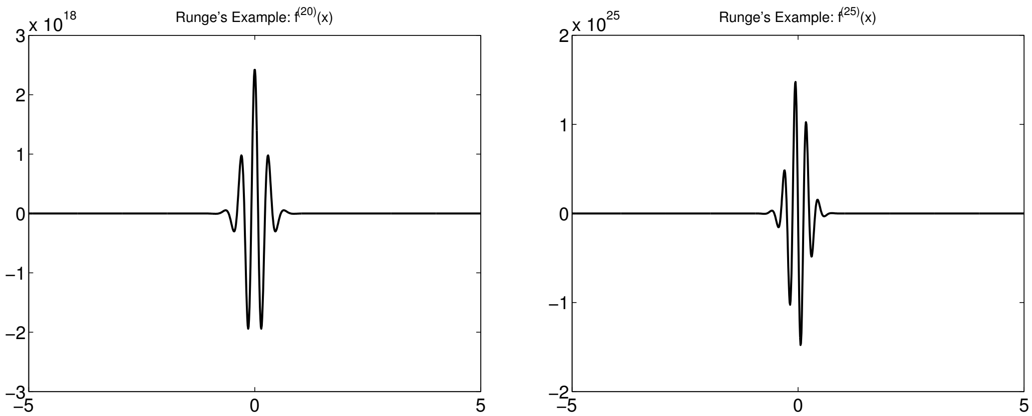
At certain points on $[-5,5]$, $f^{(n+1)}$ blows up more rapidly than $(n+1)!$, and the interpolation bound (11.1) suggests that p_n *will not* converge to f on $[-5,5]$ as n gets large. Not only does p_n fail to converge to f; the error between certain interpolation points gets enormous as n increases.

The following code uses MATLAB's Symbolic Toolbox to compute higher derivatives of the Runge function. Several of the resulting plots follow.[§] *Note how the scale on the vertical axis changes from plot to plot!*

```
% rungederiv.m
% routine to plot derivatives of Runge's example, f(x) = 1/(1+x^2) on [-5,5]
 figure(1), clf, set(gca,'fontsize',18)
 for j=0:25
    syms x
    fj = vectorize(diff(1/(x^2+1),j));              % compute derivative (Symbolic Toolbox)
    x = linspace(-5,5,1000); fjx = eval(fj);        % evaluate on a grid of points
    plot(x,fjx,'b-','linewidth',2);                 % plot derivative
    title(sprintf('Runge''s Example: f^{(%d)}(x)',j),'fontsize',14)
    input(' ')
 end
```



[§]Not all versions of MATLAB have the Symbolic Toolbox, but you should be able to run this code on any Student Edition, on Owlnet, and on the computers in the CAAM Department.



Some improvement can be made by a careful selection of the interpolation points $\{x_0\}$. In fact, if one interpolates Runge's example, $f(x) = 1/(1 + x^2)$, at the *Chebyshev points* for $[-5, 5]$,

$$x_j = 5\cos\Big(\frac{j\pi}{n}\Big), \qquad j = 0, \ldots, n,$$

then the interpolant will converge!

As a general rule, interpolation at Chebyshev points is greatly preferred over interpolation at uniformly spaced points for reasons we shall understand in a few lectures. However, even this set is not perfect: there exist functions for which the interpolants at Chebyshev points do not converge. Examples to the effect were constructed by Marcinkiewicz and Grunwald in the 1930s. We close with two results of a more general nature.¶ We require some general notation to describe a family of interpolation points that can change as the polynomial degree increases. Toward this end, let $\{x_j^{[n]}\}_{j=0}^n$ denote the set of interpolation points used to construct the degree-n interpolant. As we are concerned here with the behavior of interpolants as $n \to \infty$, so we will speak of the *system of interpolation points* $\{\{x_j^{[n]}\}_{j=0}^n\}_{n=0}^\infty$.

Our first result is bad news.

Theorem (Faber's Theorem). Let $\{\{x_j^{[n]}\}_{j=0}^n\}_{n=0}^\infty$ be any system of interpolation points with $x_j^{[n]} \in [a, b]$ and $x_j^{[n]} \neq x_\ell^{[n]}$ for $j \neq \ell$ (i.e., distinct interpolation points for each polynomial degree). Then there exists some function $f \in C[a, b]$ such that the polynomials p_n that interpolate f at $\{x_j^{[n]}\}_{j=0}^n$ do not converge uniformly to f in $[a, b]$ as $n \to \infty$.

The good news is that there always exists a suitable set of interpolation points for any given $f \in C[a, b]$.

Theorem (Marcinkiewicz's Theorem). Given any $f \in C[a, b]$, there exist a system of interpolation points with $x_j^{[n]} \in [a, b]$ such that the polynomials p_n that interpolate f at $\{x_j^{[n]}\}_{j=0}^n$ converge uniformly to f in $[a, b]$ as $n \to \infty$.

¶An excellent exposition of these points is given in volume 3 of I. P. Natanson, *Constructive Function Theory*, Ungar, New York, 1965.