CHAPTER 1

Lagrange interpolation

1. The Lagrange interpolating polynomial

Given finitely many distinct $x_0, x_1, \ldots, x_n$ real numbers (called the abscissas), and for each a corresponding value $y_i$, the problem of Lagrange interpolation is to find a polynomial function $p(x)$ of degree at most $n$ (one less than the number of abscissas and values give) and such that

$$p(x_i) = y_i, \quad i = 0, 1 \ldots n.$$ 

For example, given the US census data

<table>
<thead>
<tr>
<th>Year</th>
<th>Population</th>
</tr>
</thead>
<tbody>
<tr>
<td>1980</td>
<td>226,540,825</td>
</tr>
<tr>
<td>1990</td>
<td>248,709,873</td>
</tr>
<tr>
<td>2000</td>
<td>281,421,906</td>
</tr>
<tr>
<td>2010</td>
<td>308,745,538</td>
</tr>
</tbody>
</table>

can we find a cubic polynomial $p(x)$ (more precisely, a polynomial of degree 3 or less) such that $p(1980) = 226,540,825$, etc. Such a polynomial can be used for many purposes. For example if we want to estimate the population in 1995, we could evaluate $p(1995)$. Or we could estimate the rate of growth of the population in 1995 as $p'(1995)$.

First let us show that, for any distinct abscissas and values, the Lagrange interpolation problem has a unique solution. One approach is write the sought after polynomial in the usual way as a sum of monomials:

$$p(x) = \sum_{i=0}^{n} a_i x^i,$$

with the coefficients $a_i$ to be determined. This gives us a system of $n + 1$ linear equations in $n + 1$ unknowns:

$$\begin{pmatrix}
 1 & x_0 & x_0^2 & \cdots & x_0^n \\
 1 & x_1 & x_1^2 & \cdots & x_1^n \\
 \vdots & \vdots & \vdots & \ddots & \vdots \\
 1 & x_n & x_n^2 & \cdots & x_n^n
\end{pmatrix} \begin{pmatrix}
a_0 \\
a_1 \\
\vdots \\
 a_n
\end{pmatrix} = \begin{pmatrix}
y_0 \\
y_1 \\
\vdots \\
y_n
\end{pmatrix}$$

Thus the question of existence and uniqueness comes down to whether this $(n + 1) \times (n + 1)$ matrix is invertible. To see that it is, we need to show that the only solution to (1.2) when all the $y_i$ are 0 is the trivial solution. Indeed, any solution gives us coefficients $a_i$ of a polynomial of degree at most $n$ which has $n + 1$ roots, and, from algebra, we know that this means the polynomial must vanish. The conclusion is that the Vandermonde matrix is nonsingular, and so the Lagrange interpolation problem always has a unique solution.
1. LAGRANGE INTERPOLATION

To actually compute and evaluate the interpolating polynomial, we could use this approach: (1) form the Vandermonde matrix, (2) solve the linear system (1.2), (3) evaluate the polynomial from (1.1). However, there are better ways to do this, both in terms of efficiency of stability.

Lagrange’s form. Suppose we form the \( n+1 \) polynomials

\[
L_j(x) = \frac{n}{\prod_{m=0, m\neq j}^{n} (x - x_m)} \prod_{m=0}^{n} (x_j - x_m), \quad j = 0, 1, \ldots, n.
\]

Then

\[
L_j(x_i) = \begin{cases} 
1, & i = j, \\
0, & \text{else}
\end{cases}
\]

Therefore

\[
p(x) = \sum_{j=0}^{n} y_j L_j(x)
\]

solves the Lagrange interpolation problem. This formula is known as Lagrange’s form of the interpolating polynomial. The polynomial it determines is the same as that determined by solving the Vandermonde system but there is no need to solve a linear system to compute it. With careful programming the Lagrange form can be useful and practical, although we shall generally prefer yet another formula, Newton’s form of the interpolating polynomial, which we shall introduce below.

1.1. Error formula. Now suppose that there is a function \( f \), probably not a polynomial, defined on an interval containing all the abscissas and such that the values are just the values of \( f \): \( y_i = f(x_i) \). We then consider the difference

\[
E(x) = f(x) - p(x).
\]

To get a handle on this error, we fix one point \( x \) and consider the function

\[
G(t) = [f(x) - p(x)](t - x_0) \cdots (t - x_n) - [f(t) - p(t)](x - x_0) \cdots (x - x_n).
\]

Clearly \( G \) vanishes when \( t \) is one of the \( (n+1) \) points \( x_j \). We have also constructed it so that \( G(t) = 0 \) when \( t = x \). Thus the function \( G \) has \( n + 2 \) roots. We shall suppose that the function \( f \) is differentiable as many times as needed. Then, applying Rolle’s theorem or the Mean Value Theorem, we conclude that the derivative \( G' \) has at least \( n + 1 \) roots. The same reasoning then shows that \( G'' \) has \( n \) roots and, finally, that \( G^{(n+1)}(\xi) = 0 \) for some \( \xi \) in the smallest interval containing the \( x_j \) and \( x \). Now

\[
p^{(n+1)}(\xi) = 0, \quad d^{n+1}[(t - x_0) \cdots (t - x_n)]/dx^{n+1} = (n + 1)!,
\]

so

\[
f(x) - p(x) = \frac{1}{(n + 1)!} f^{(n+1)}(\xi)(x - x_0) \cdots (x - x_n). \tag{1.3}
\]
Figure 1.1. The error product for 17 equally spaced abscissas from 0 to 4.

Theorem 1.1 (Error formula for Lagrange interpolation). Suppose $f \in C^{n+1}(I)$ on some interval $I$ containing $n+1$ distinct abscissas $x_i$ and another point $x$. Let $p$ be the Lagrange interpolating polynomial interpolating the values $f(x_i)$ at $x_i$. Then, for any point there is number $\xi$ in the least interval containing the $x_i$ and $x$, such that (1.3) holds.

It is not easy to get a good quantitative bound on the error from this formula, even for quite simple functions $f(x)$ and simple point sets $x_i$. This is because, the higher derivatives $f^{n+1}$ generally become very complicated. But we can learn a lot by studying the final factor, $\omega(x) = (x-x_0) \cdots (x-x_n)$, often called the error product associated to the abscissas.

For example, Figure 1.1 shows the error product for 17 equally spaced abscissas from 0 to 4. The point to notice is that the error is many orders of magnitude larger near the ends of the interval than near the center. This illustrates a big problem with Lagrange interpolation using equally spaced abscissas and it is a major reason why it is rarely used for large number of points (more than perhaps 8, say). For larger $n$ it tends to have large errors near the interval end points. In particular, if we fix the interval and the function, say a very nice smooth function, and add more and more equally spaced points to obtain Lagrange interpolating polynomials of higher and higher degree, then it is usually not the case that the resulting that the polynomials converge to the function. To the contrary, the polynomials often oscillate wildly near the interval ends. A famous example, due to Runge, is $f(x) = 1/(1 + x^2)$ on $[-5, 5]$. It is shown in Figure 1.2 for $n = 12$.

2. Best Polynomial approximation

The results of the last section might suggest that the whole notion of approximating functions by polynomials is doomed. This is not the case. Any continuous function on a closed bounded interval can be approximated arbitrarily accurately by a polynomial. This result is called the Weierstrass Approximation Theorem. To state it carefully, we introduce the $L^\infty$ norm for functions on an interval $I$:

$$\|f\|_{\infty} := \sup_{x \in I} |f(x)|.$$
Theorem 1.2 (Weierstrass Approximation). Let \( f \in C(I) \) for a closed bounded interval \( I \) and let \( \epsilon = 0 \). Then there exists a polynomial \( p \) such that \( \| f - p \|_\infty \leq \epsilon \).

There are many proofs of the Weierstrass Approximation Theorem, but none of them are easy, and we shall not present a proof here. Now we consider the speed of convergence. Let \( P_n \) denote the space of polynomials of degree at most \( n \). For \( f \in C(I) \), let

\[
E_n(f) = \inf_{p \in P_n} \| f - p \|_\infty
\]

denote the error in the best possible approximation of \( f \) by polynomials of degree \( n \). Although it is not crucial at this point, we remark that the best approximation is achieved: for each \( n \) there exists a polynomial \( p_n(f) \in P_n \) for which \( \| f - p_n(f) \|_\infty = E_n(F) \) and, in fact, it is even unique. In any case, by the Weierstrass theorem, we have

\[
\lim_{n \to \infty} E_n(f) = 0.
\]

But how fast is this convergence? This question was answered by Dunham Jackson. Jackson’s theorem shows that if \( f \in C^k(I) \), then \( E_n(f) \) decreases as \( O(n^{-k}) \). Thus for smooth functions, the convergence is very fast indeed.

Theorem 1.3 (Jackson’s Theorem). Let \( k \) be a positive integer. Then there exists a number \( c \) (depending on \( k \)) such that

\[
\inf_{p \in P_n} \| f - p \|_\infty \leq cn^{-k}\| f^{(k)} \|_\infty.
\]

If \( f \in C^\infty(I) \) then Jackson’s theorem implies that \( E_n(f) \) decreases to zero faster than any power of \( 1/n \). In fact, if \( f \) is real analytic (meaning it is \( C^\infty \) and its Taylor expansion converges at each point), then one can prove that \( E_n(f) = O(\delta^n) \) for some positive \( \delta < 1 \), i.e., the rate of convergence is exponential.

The theorems of Weierstrass and Jackson provide two aspects of the best approximation by polynomials of degree \( n \). On the one hand, they converge for any continuous function \( f \). On the other, if \( f \) is smooth, they converge quickly. These are very desirable properties, but, unfortunately, it is not so easy to compute the best approximation. In the next sections
3. Interpolation at the Chebyshev points

From the previous sections we find that any nice function there exists a sequence of polynomials that approximate it well and quickly, but we have no easy way of computing these polynomials. We can compute the Lagrange interpolating polynomials easily, but these do not generally approximate well as the number of points is increased, at least if we use equally spaced abscissas. A way around this situation is to use Lagrange interpolation with a better choice of points.

Recalling the error formula for Lagrange interpolation we see that a way to reduce the error is to choose the interpolation points $x_i$ so as to decrease error product $\omega(x) = (x - x_0) \cdots (x - x_n)$. Thus we are led to the problem of finding $x_0 < \cdots < x_n$ which minimize $\|\omega\|_{\infty}$. In fact, we can solve this problem in closed form. For simplicity we take the interval to be $[-1, 1]$. On this interval define $T_n(x) = \cos(n \arccos x)$. Then, using the additional formula for cosine, we find

$$T_0(x) = 1, \quad T_1(x) = x, \quad T_{n+1}(x) = 2xT_n(x) - T_{n-1}(x), n = 1, 2, \ldots.$$ 

Thus $T_n(x)$ is a polynomial of exact degree $n$ with leading coefficient $2^{n-1}$.

Now let

$$x_i^n = \cos[(2i + 1)\pi/(2n + 2)], \quad i = 0, 1, \ldots, n.$$ 

Then it is easy to see that $1 > x_0^n > x_1^n > \cdots > x_n^n > -1$ and that these are precisely the $n + 1$ zeros of $T_{n+1}$. These are called the $n + 1$ Chebyshev points on $[-1, 1]$. The definition is illustrated for $n = 8$ in Figure 1.3. The next theorem shows that the Chebyshev points minimize the error product in $L^\infty$.

**Theorem 1.4.** For $n \geq 0$, let $x_0, x_1, \ldots, x_n \in \mathbb{R}$ by the Chebyshev points on the interval $[-1, 1]$ (the roots on $T_{n+1}$) and let $\omega(x) = (x - x_0) \cdots (x - x_n)$ be the corresponding error product. Then

$$\|\omega\|_{\infty} = 2^{-n}.$$ 

Moreover if $\tilde{x}_i$ are any distinct points in $[-1, 1]$ and $\tilde{\omega}$ the corresponding error product, then

$$\|	ilde{\omega}\|_{\infty} \geq 2^{-n}.$$
Proof. First assume that the \( x_i \) are the \( n+1 \) Chebyshev points. Then \( \omega \) and \( T_{n+1} \) are two polynomials of degree \( n+1 \) with the same roots. Comparing their leading coefficients we see that \( \omega(x) = 2^{-n}T_{n+1}(x) = 2^{-n} \cos(n \arccos x) \). The first statement follows immediately. Note also that for this choice of points, \( |\omega(x)| \) achieves its maximum value of \( 2^{-n} \) at \( n+2 \) distinct points in \([-1, 1]\), namely at \( \cos[j\pi/(n+1)], j = 0, \ldots, n+1 \), and that the sign of \( \omega(x) \) alternates at these points.

Now suppose that some other points \( \tilde{x}_i \) are given and set \( \tilde{\omega}(x) = (x - \tilde{x}_0) \cdots (x - \tilde{x}_n) \). If \( \|\tilde{\omega}\|_\infty \) were strictly less than \( 2^{-n} \) on \([-1, 1]\), then the difference \( \omega(x) - \tilde{\omega}(x) \) would alternate sign at the \( n+2 \) points \( \cos[j\pi/(n+1)] \) and so would have at least \( n+1 \) zeros. But it is a polynomial of degree at most \( n \) (since the leading terms cancel), and so must vanish identically, a contradiction. \( \square \)

Of course we can easily compute the Chebyshev points on any interval \([a, b]\) by linear scaling. On \([a, b]\) we use the points

\[
x_i = \frac{a + b}{2} + \frac{b - a}{2} \cos[(2i + 1)\pi/(2n + 2)], \quad i = 0, \ldots n.
\]

These give the smallest error product when measured in \( L^\infty([a, b]) \). For example, compare the case \( n = 16 \) on the interval \([0, 4]\) shown in Figure 1.4 with the error product using the same number of equally space points shown in Figure 1.1. If we recompute the degree 12 interpolant to Runge’s example function using Chebyshev points, we get the Figure 1.5, starkly contrasting with Figure 1.2.

We now summarize the main results for approximation of a function by its interpolant at the Chebyshev points. To be precise we recall the definition of a Hölder continuous function. This means that there exists a constant \( C \) and a number \( \alpha > 0 \) such that \( |f(x) - f(y)| \leq C|x - y|^{\alpha} \) for all \( x, y \in I \). It includes all \( C^1 \) functions, but also continuous function except for a finite number of kinks (like \( f(x) = |x| \) on \([-1, 1]\)) and even worse functions, like \( f(x) = |x|^{1/3} \) which has a cusp.

**Theorem 1.5.** Let \( f \) be a Hölder continuous function on \( I \) and let \( p_n \) be the Lagrange interpolant to \( f \) at \( n+1 \) Chebyshev points on \( I \). Then \( \lim_{n \to \infty} \|f - p_n\|_\infty = 0 \).
This tells us that interpolation at the Chebyshev points converges for almost any continuous function $f$. However it is not true that it converges for every continuous function (unlike the best polynomial approximation, which does, thanks to the Weierstrass Approximation Theorem).

More importantly, if the function $f \in C^k$ for some $k \geq 1$, we get a rate of convergence which is nearly the same as that given by the Jackson theorem.

**Theorem 1.6.** Let $f \in C^k(I)$ for some $k \geq 1$ and let $p_n$ denote the interpolant at the Chebyshev points. Then there exists $c$ such that

$$\|f - p\|_\infty \leq cn^{-k} \log n \|f^{(k)}\|_\infty.$$ 

Again, if $f$ is $C^\infty$ the convergence is faster than any power of $1/n$ and if $f$ is analytic, then the convergence is exponential.

### 4. Best $L^2$ approximation

In this section we stick to the interval $I = [-1, 1]$ (we can always use linear scaling to change the interval), and consider the question of best $L^2$ approximation by a polynomial. That is, give a function $f$ on $I$, we want to find a polynomial $p_n \in P_n(I)$ such that $\|f - p\|_2$ is minimal. This problem is much easier for the $L^2$ norm than for the $L^\infty$ norm, because the $L^2$ norm is associated with an inner product. Define the $L^2$ inner product of two functions $f$ and $g$ to be

$$(f, g)_2 = \int_{-1}^{1} fg \, dx,$$

so

$$\|f\|_2 = \left(\int_{-1}^{1} f^2 \, dx\right)^{1/2}.$$ 

(For most of this section, we will drop the subscript 2 from the norm and inner product.)
Theorem 1.7. Let \( f \in L^2(I) \) and let \( p \in \mathcal{P}_n \). Then \( p \) is the \( L^2 \) best approximation of \( f \) in \( \mathcal{P}_n \) if and only if
\[
\int_{-1}^{1} (f - p)q \, dx = 0, \quad q \in \mathcal{P}_n.
\]

Proof. Suppose \( q \) is any element of \( \mathcal{P}_n \). Then
\[
F(t) = \|f - (p + tq)\|^2 = \|f - p\|^2 + 2t(f - p, q) + t^2\|q\|^2.
\]
Then \( F(t) \) is a quadratic polynomial in \( t \) and
\[
F'(t) = 2(f - p, q) + 2t\|q\|^2.
\]
The function \( p \) is a minimizer of \( \|f - p\|^2 \) if and only if \( F \) has a minimum at \( t = 0 \) for all \( q \), which holds if and only if the coefficient \( (f - p, q) \) vanishes for all \( q \). \( \square \)

The equations (1.4) states that the minimizer \( p \) is the \( L^2 \) orthogonal projection of \( f \) onto the subspace \( \mathcal{P}_n \). The proof also points to a simple way to compute the projection. If we choose a convenient basis \( L_0, L_1, \ldots, L_n \) of \( \mathcal{P}_n \) and write the minimizer
\[
p = \sum_{m=0}^{n} a_m L_m
\]
with coefficients \( a_m \) to be determined, we find that
\[
\begin{pmatrix}
(L_0, L_0) & \cdots & (L_n, L_0) \\
\vdots & \ddots & \vdots \\
(L_n, L_0) & \cdots & (L_n, L_n)
\end{pmatrix}
\begin{pmatrix}
a_0 \\
a_1 \\
\vdots \\
a_n
\end{pmatrix} =
\begin{pmatrix}
(f, L_0) \\
(f, L_1) \\
\vdots \\
(f, L_n)
\end{pmatrix}
\]
Thus, we first compute the matrix (called the Gram matrix for the basis), then compute the integrals of \( f \) with the basis functions, and then solve the \((n + 1) \times (n + 1)\) linear system to find the coefficients \( a_m \). The polynomial we seek is then given by the expansion (1.5).

This approach is quite practical. However, as for Lagrange interpolation, it is improved by choosing a basis suited for the problem. If the basis is orthogonal, meaning \((L_i, L_j) = 0 \) if \( i \neq j \), then the Gram matrix is diagonal, and so solving it is trivial:
\[
p = \sum_{j=0}^{n} \frac{(f, L_j)}{(L_j, L_j)} L_j.
\]
If the basis is orthonormal, we can drop the denominators as well.

It turns out that it is quite easy to construct an orthogonal basis of \( \mathcal{P}_n \) without solving a linear system. In a word, we may start with the monomial basis: 1, \( x \), \( x^2 \), \ldots, and then apply the Gram-Schmidt orthogonalization process. In fact, for polynomials the process simplifies considerably. To see how it works, suppose we have polynomials \( p_0, p_1, \ldots, p_n \) which are mutually orthogonal, and with the degree of \( p_m = m \). Suppose also that \( p_m \) is an even function for \( m \) even and an odd function for \( m \) odd. Consider defining
\[
p_{n+1} = xp_n + \alpha_n p_{n-1}.
\]
This clearly defines a polynomial of degree $n + 1$ and we consider how to choose $\alpha_n$ such the $p_{n+1}$ is orthogonal to everything in $P_n$. We obtain a great deal of orthogonality immediately: if $q \in P_{n-2}$, then $xq \in P_{n-1}$, and so

$$(p_{n+1}, q) = (p_n, xq) + \alpha_n(p_{n-1}, q) = 0.$$ 

Also $p_{n+1}$ is orthogonal to $p_n$, just because one is odd and one is even. It remains only to insure that $(p_{n+1}, p_{n-1}) = 0$, which determines $\alpha_n$ as

$$\alpha_n = -\frac{(xp_n, p_{n-1})}{\|p_{n-1}\|^2}.$$

To initialize this method we define $p_0 = 1$, $p_1 = x$, and then set

$$p_{n+1}(x) = xp_n - \frac{(xp_n, p_{n-1})}{\|p_{n-1}\|^2} p_{n-1}(x), \quad n = 1, 2, \ldots.$$ (1.6)

The polynomials so defined are called the Legendre polynomials. As we have defined them they are monic:

$$p_n(x) = x^n + \text{lower order terms}.$$ 

It is more common to normalize the Legendre polynomials to take the value 1 at $x = 1$ and to denote them by $P_n$. It can then be shown that the recursion becomes even a little simpler:

$$P_{n+1}(x) = \frac{2n+1}{n+1}xP_n(x) - \frac{n}{n+1}P_{n-1}(x).$$ (1.7)

(Compare to the recursion defining the Chebyshev polynomials. The form is the same and the coefficients only slightly different: $(2n+1)/(n+1)$ instead of 2 and $-n/(n+1)$ instead of $-1$.) To derive (1.7) from (1.6) requires some algebraic computation, which we skip.

The Legendre polynomials are then:

$$P_0(x) = 1, \quad P_1(x) = x, \quad P_2(x) = \frac{3}{2}x^2 - \frac{1}{2}, \quad P_3(x) = \frac{5}{2}x^3 - \frac{3}{2}x, \quad P_4(x) = \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}, \ldots.$$ 

See their graphs in Fig. 1.6.

Notice that the polynomial $P_n$ has all $n + 1$ of its roots in the open interval $(-1, 1)$. In fact, if it changed sign only at the points $x_0 < \ldots < x_m$ in $(-1, 1)$ with $m < n$, then the polynomial $(x - x_0) \cdots (x - x_m)$ would be orthogonal to $P_n$, but their product would never change sign, which is impossible.

In summary, to compute the best $L^2$ approximation, we first compute (or look up) the Legendre polynomials of degree up to $n$, and then use the formula

$$p = \sum_{j=0}^n c_j(f, P_j)P_j.$$ 

where

$$c_j = \frac{1}{\|P_j\|^2} = \frac{2n + 1}{2}.$$

(The last equality is some more algebra that we skip.)
Finally, we note that the error is easy to bound if we use Jackson’s theorem. For any function $\|g\|_2 \leq \sqrt{2}\|g\|_\infty$. If $p_n$ is the best $L^2$ approximation of $f$ in $\mathcal{P}_n$, and $q$ is any function in $\mathcal{P}_n$, we then have

$$\|f - p_n\|_2 \leq \|f - q\|_2 \leq \sqrt{2}\|f - q\|_\infty \leq Cn^{-k}\|f^{(k)}\|_\infty.$$