Towards best polynomial approximation

The absolute error of Lagrange interpolation, as we now know, is given by the expression:

$$|R(x)| = \frac{|f^{(n)}(\xi(x))|}{n!} \prod_{i=1}^{n} |x - x_i|. \tag{1}$$

We can regard (1) as a product of two factors: the absolute value of the derivative divided by the factorial, and the absolute value of the nodal polynomial. The first of these factors depends on the function that is being interpolated and is, generally, beyond our control. At best we may have an upper bound: $|f^{(n)}(x)| \leq M$, although even that cannot be taken for granted. In contrast, the absolute value of the nodal polynomial is independent of the function and can be controlled, to some extent, by moving the nodes. This raises the following question:

> If we are free to choose nodes, where should the nodes be located so that the contribution of $\prod_{i=1}^{n} |x - x_i|$ to the error of interpolation is as small as possible?

Specifically, if we interpolate on some interval $[a, b]$ we would like to find a set of nodes that minimizes the quantity:

$$\max_{a \leq x \leq b} \prod_{i=1}^{n} |x - x_i|.$$

Such a choice may be considered "best" in the sense that it minimizes the error (1).

The concept of norm

At this point it will be helpful to introduce some terminology that will help us avoid awkward phrases like "a set of nodes that minimizes the quantity." That quantity, actually, is the so-called *infinity norm* of the nodal polynomial which is just one example of several different norms that we will encounter in the future.

Think of 'norm' as a generalization of 'length.' There are many ways to generalize length and, consequently, there are many norms. Some of the norms used by mathematicians are very strange and do not resemble length at all. We therefore precede the formal definition of norm with a few remarks that should help with the abstraction.

1. A norm is an attribute of a vector. So, when we say "infinity norm of the nodal polynomial" we refer to the "length" of the nodal polynomial which we regard as a vector. You will recall from our earlier discussions of basic linear algebra that vectors make very little sense individually. In order to be useful they must be organized into collectives called *vector spaces*. It is therefore more accurate to say that norm is an attribute of a *vector space*. In fact, advanced texts on functional analysis usually define *normed vector spaces* instead of defining norms.

2. Before reading the formal definition of the norm, think of the various statements which can be made about the usual length of a vector in $\mathbb{R}^n$. For instance, length must be nonnegative: that means that norm must be nonnegative. Furthermore, the only vector whose length is zero is the zero vector. Hence zero norm must imply that the vector is actually the zero vector.

3. A vector space is, loosely speaking, a collection of objects which can be added and scaled (like vectors in $\mathbb{R}^n$.) That means that there must be some interplay between the norm and the vector space operations. We should be able to say something about the norm of a scaled vector and the norm of a sum of vectors.

We are now ready for formalism. Let V be a vector space.

Definition 2. *The norm $\|\cdot\|$ on a vector space V is a function taking vectors in V into real numbers. Furthermore, if v and w are vectors in V and a is a scalar then the following must be true:*

1. *The norm must be nonnegative: $\|v\| \geq 0$. Furthermore, if $\|v\| = 0$ then $v = 0$ (and conversely).*

2. *The norm must have a scaling property: $\|a\, v\| = |a|\, \|v\|$, where $|a|$ is the absolute value of the scalar.*

3. *The triangle inequality must hold: $\|v + w\| \leq \|v\| + \|w\|$.*

You may wonder why the last property in Definition 2 is called the triangle inequality. This is explained by Figure 3: the length of a side in a triangle is

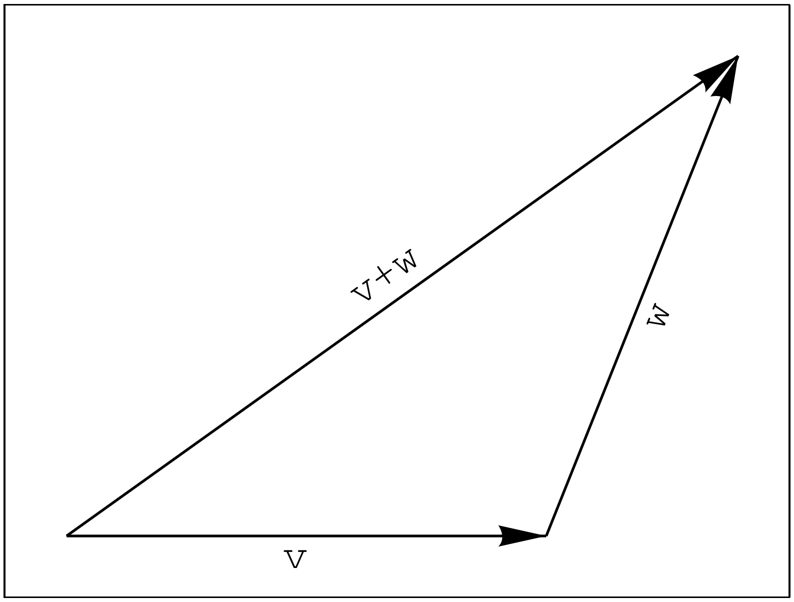

Figure 3: Triangle inequality

smaller than (or, in the degenerate case, equal to) the sum of the lengths of the other two sides.

Let v be a vector in $\mathbb{R}^n$ with components $v_1, \ldots, v_n$ and let p be a positive number greater or equal to one. The p-norm of v is the following expression:

$$\|v\|_p = \left(\sum_{i=1}^{n} |v_i|^p \right)^{\frac{1}{p}}$$

Among the p-norms the following three cases are the most important ones:

$p = 1$: $\|v\|_1 = \sum_{i=1}^{n} |v_i|$

$p = 2$: $\|v\|_2 = \sqrt{\sum_{i=1}^{n} v_i^2}$

$p = \infty$: $\|v\|_\infty = \max\{|v_1|, \ldots, |v_n|\}$

As an exercise, convince yourself that the given expression for the ∞-norm is, indeed, the limit of the p-norm: $\lim_{p \to \infty} \|v\|_p$.

If vectors are functions, as it is usually the case in Numerical Analysis, the p-norms are defined using integration. For instance, if f is a function on $[a, b]$, its p-norm is given by:

$$\|f\|_p = \left(\int_a^b |f(x)|^p \, dx \right)^{\frac{1}{p}}$$

In particular, the most important functional p-norms are:

$p = 1$**:** $\|f\|_1 = \int_a^b |f(x)| \, dx$

$p = 2$**:** $\|f\|_2 = \sqrt{\int_a^b f^2(x) \, dx}$

$p = \infty$**:** $\|f\|_\infty = \sup_{a \le x \le b} |f(x)|$

Notice that we had to use the supremum to define the ∞-norm because a maximum may fail to exist if the function is discontinuous, for example. Because of that the ∞-norm is often called the supremum-, or sup-norm for short.

We are now ready to return to the original question that prompted the discussion of norms. Evidently, the best nodes for interpolation are the ones that minimize the ∞-norm of the nodal polynomial. We therefore would like to find a nodal polynomial with the smallest ∞-norm.

Some quick numerical results

Since the classical solution due to Chebyshev is for the interval $[-1, 1]$, we will use that interval. It is clear from symmetry that for one node the nodal polynomial should be, simply, $p_1 = x$. To confirm that, let us take $x_0 \in [-1, 1]$. Then, since $x - x_0$ does not have extrema within $(-1, 1)$:

$$\|x - x_0\|_\infty = \max\{|-1 - x_0|, |1 - x_0|\} = \begin{cases} |1 + x_0|, & \text{if } x_0 \ge 0; \\ |1 - x_0|, & \text{if } x_0 < 0. \end{cases}$$

Notice that

$$\|x - x_0\|_\infty \ge 1$$

with equality for $x_0 = 0$. Therefore

$$\min_{-1 \le x_0 \le 1} \|x - x_0\|_\infty = 1$$

and $p_1 = x$ as we expected.

We could compute p_2 using simple Calculus, however, it will be instructive to compute it numerically.

```
function chebyshev
% Minimize the sup-norm of a nodal polynomial on [-1,1]
% The solution is given by monic Chebyshev polynomials
% of the first kind.

N = 2;                              % degree of the polynomial
r0 = -1 + 2*(1:N)/(N+1);            % approximate roots
r = fminsearch(@supnorm,r0);        % minimize sup-norm
p = poly(r);                        % find polynomial from roots

% Plot the absolute value of p, the initial guess for its
% roots, and optimized roots.

x = linspace(-1,1,1024);
plot(x,abs(polyval(p,x)), 'b-',r,0,'ro',r0,0,'bo')
end

function f = supnorm(r)
p = poly(r);
x = roots(polyder(p));
f = max(abs(polyval(p,[-1;x;1])));
end
```

The code computes p_2 as

```
1.0000   -0.0000   -0.5000
```

which is to say: $p_2 = x^2 - \frac{1}{2}$. The plot of the absolute value of p_2 is shown in Figure 4. Notice that $|p_2|$ attains its maximum value of $\|p_2\|_\infty = \frac{1}{2}$ at three points: the endpoints of the interval and $x = 0$.

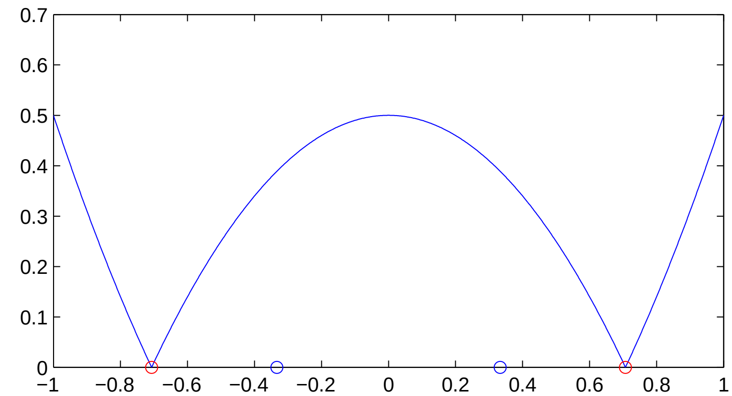

Figure 4: Absolute value of p_2

Running `chebyshev` with $N = 3$ produces

```
1.0000     0.0055    -0.7473    -0.0055
```

which suggests that $p_3 = x^3 - \frac{3}{4}x$. Figure 5 shows that $|p_3|$ attains its maximum $\|p_3\|_\infty = \frac{1}{4}$ at four places.

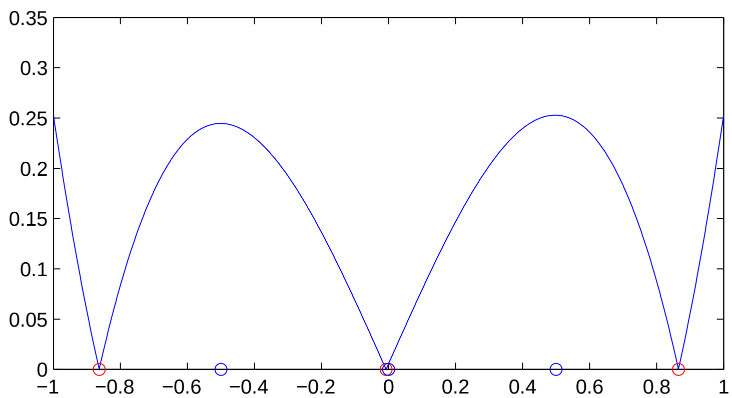

Figure 5: Absolute value of p_3

Unfortunately, the code fails to find an accurate approximation to p_4. The problem is with `fminsearch`: it gets stuck at a local minimum. There is a short-term fix for that: we can use symmetry. It is clear that the roots of p_4 should come in two pairs which are symmetric with respect to the origin. We can therefore minimize in two dimensions rather than four—over the positive roots. If that strategy is implemented (exercise) the result is:

```
1.0000          0    -1.0000    0.0000    0.1250
```

Therefore

$$p_4 = x^4 - x^2 + \frac{1}{8}$$

and the plot of its absolute value is given in Figure 6. Again, notice how $|p_4|$ attains its maximum $\|p_4\|_\infty = \frac{1}{8}$ exactly five times.

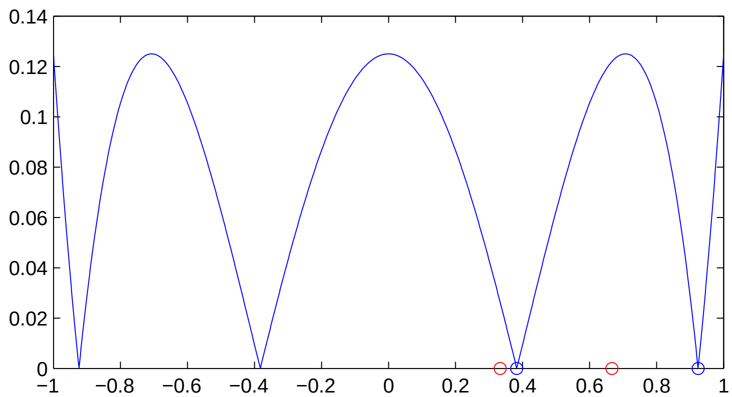

Figure 6: Absolute value of p_4

Summarizing our results, the first four nodal polynomials with minimum ∞-norms on $[-1, 1]$ are:

$$x, \quad x^2 - \frac{1}{2}, \quad x^3 - \frac{3}{4}x, \quad x^4 - x^2 + \frac{1}{8}$$

From the plots of absolute values it is clear that $\|p_n\|_\infty = \frac{1}{2^{n-1}}$. Furthermore, $|p_n|$ attains its maximal value $\|p_n\|_\infty$ exactly $(n+1)$ times. This important property is called *equioscillation*.

Unfortunately, it is difficult to see the pattern in the coefficients of p_n. Therefore let us look at the corresponding nodes:

p_n	nodes
p_1	0
p_2	$-\frac{1}{\sqrt{2}}, +\frac{1}{\sqrt{2}}$
p_3	$-\frac{\sqrt{3}}{2}, 0, +\frac{\sqrt{3}}{2}$
p_4	$-\frac{\sqrt{2+\sqrt{2}}}{2}, -\frac{\sqrt{2-\sqrt{2}}}{2}, +\frac{\sqrt{2-\sqrt{2}}}{2}, +\frac{\sqrt{2+\sqrt{2}}}{2}$

Table 1: Nodes of Chebyshev polynomials

Notice how the nodes of p_2 and p_3 are reminiscent of standard values of trigonometric functions. In fact, the two nodes of p_2 can be written as:

$$\cos\left(\frac{\pi}{4}\right), \quad \cos\left(\frac{3\,\pi}{4}\right)$$

while the nodes of p_3 can be written as:

$$\cos\left(\frac{\pi}{6}\right), \quad \cos\left(\frac{3\,\pi}{6}\right), \quad \cos\left(\frac{5\,\pi}{6}\right)$$

It is now easy to guess that the nodes of p_4 are

$$\cos\left(\frac{\pi}{8}\right), \quad \cos\left(\frac{3\,\pi}{8}\right), \quad \cos\left(\frac{5\,\pi}{8}\right), \quad \cos\left(\frac{7\,\pi}{8}\right)$$

and, in general, the nodes of p_n are:

$$\left\{\cos\left(\frac{\pi\,(2\,k-1)}{2\,n}\right), \quad k = 1, \ldots, n\right\}$$

Figure 7 confirms that. It shows the plot of p_5 obtained using the following commands:

```
n = 5;
p = poly(cos(pi*(1:2:2*n)/(2*n)));
x = linspace(-1,1,1024);
y = abs(polyval(p,x));
plot(x,y)
```

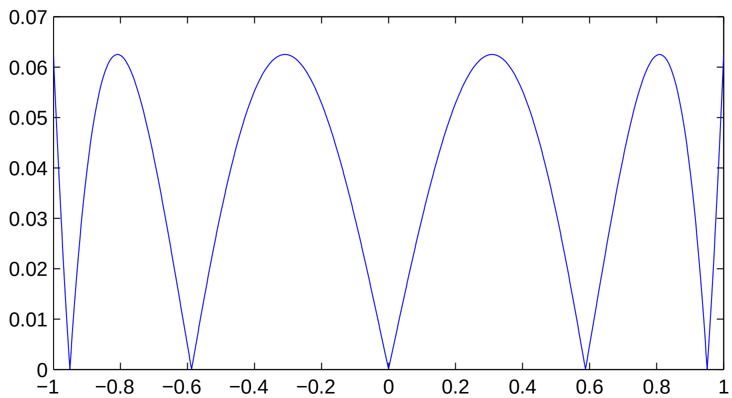

Figure 7: Absolute value of p_5

Notice how the plot attains its maximum $\frac{1}{2^{5-1}}$ five times.

Our next observation is that the nodes of p_n can be thought of as n solutions of the equation:

$$n \arccos(x) = \frac{\pi}{2}.$$

Since $\frac{\pi}{2} = \arccos(0)$, we can rewrite the equation in the form:

$$\cos(n \arccos(x)) = 0.$$

This, actually, almost gives us the polynomials we are looking for! It may not seem like it but $\cos(n \arccos(x))$ is actually a polynomial for integer n. This is obvious for $n = 1$: indeed, we can recover p_1 as $\cos(\arccos(x)) = x$. However, if $n = 2$ we can use the double angle formula

$$\cos(2\theta) = \cos^2(\theta) - \sin^2(\theta) = 2 \cos^2(\theta) - 1$$

to recover a multiple of p_2:

$$\cos(2\,\arccos(x)) = 2\,x^2 - 1 = 2\,\left(x^2 - \frac{1}{2}\right)$$

Therefore

$$p_2 = \frac{1}{2}\,\cos(2\,\arccos(x)).$$

Similarly, $\cos(3\,\arccos(x))$ can be simplified to a cubic using appropriate triple angle identities (exercise). It is easy to verify that

$$p_n = \frac{1}{2^{n-1}}\,\cos(n\,\arccos(x)).$$

The polynomials $T_n = \cos(n\,\arccos(x))$ are called *Chebyshev polynomials of the first kind*. The sequence of Chebyshev polynomials starts with the constant $T_0 = 1$.

Chebyshev's proof

We are now ready for one of the main theoretical results concerning best polynomial approximation. Recall that we are working on the interval $[-1, 1]$ and by "best approximation" we presently mean interpolation at the nodes that minimize the ∞-norm of the nodal polynomial. In order to state the classical theorem, due to Chebyshev, we need one more adjective—*monic*—which is used to describe polynomials with highest coefficient equal to 1. Any nodal polynomial is, clearly, monic.

Theorem 7. *Among all monic polynomials of degree n on the interval $[-1, 1]$ the scaled Chebyshev polynomial of the first kind*

$$p_n = \frac{1}{2^{n-1}}\,\cos(n\,\arccos(x)) = \frac{1}{2^{n-1}}\,T_n$$

has the smallest ∞-norm: $\|p_n\|_\infty = \frac{1}{2^{n-1}}$.

Proof. Suppose that the monic polynomial of degree n with the smallest ∞-norm on $[-1, 1]$ is some mystery polynomial q_n. We will show that $p_n - q_n = 0$. The key to the argument is the equioscillation property of p_n: it attains its maximum value $\frac{1}{2^{n-1}}$ exactly $(n+1)$ times at points $x_k^* = \cos\left(\frac{\pi\,k}{n}\right)$, $k = 0, \ldots, n$.

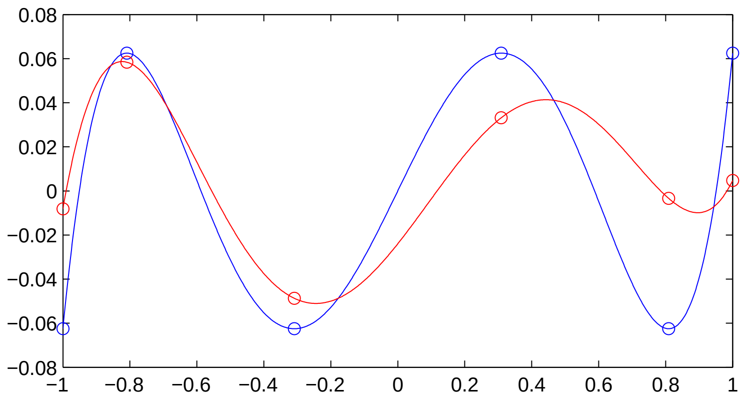

Figure 8: The blue line is the plot of p_5; the red line shows what the plot of q_5 would look like if it were distinct from p_n.

If q_n is to have smaller ∞-norm than p_n, its graph must lie within the range $[-2^{1-n}, 2^{1-n}]$ as schematically shown in Figure 8. (The red line only shows what the plot of q_n should look like–it was produced by plotting a *non*-monic polynomial). In particular, the values of q_n at the critical points of p_n should be smaller in magnitude than 2^{1-n}:

$$|q_n(x_k^*)| < |p_n(x_k^*)| = \frac{1}{2^{n-1}}, \quad k = 0, \ldots, n.$$

Think about what that says about the difference $p_n - q_n$: it must change sign in $[-1, 1]$ at least n times and therefore it must have n roots there. Yet, since both p_n and q_n are monic, the difference $p_n - q_n$ is a polynomial of degree strictly *less* than n: therefore, if it has n roots, it must be identically zero. □

Interpolation at Chebyshev nodes

Theorem 7 shows that, given a choice of nodes in $[-1, 1]$, we should pick the roots of Chebyshev polynomials of the first kind. However, it turns out that an even better system of nodes is given by the critical points of Chebyshev polynomials:

$$x_k = \cos\left(\frac{\pi\, k}{n}\right), \quad k = 0, \ldots, n.$$

These are usually called simply *Chebyshev nodes*. The accuracy of interpolation on Chebyshev nodes is about the same as for roots of Chebyshev polynomials. The advantage of Chebyshev nodes is the speed at which interpolating polynomials can be computed. Consider the following code.

```
N = 10000;
tt = cos(pi*(N:-1:0)/N);

% generate values for interpolation

f = @(t) sign(sin(100*t./(2-t)));
options = odeset('AbsTol',1e-12,'RelTol',1e-12);
[tt,yy] = ode45(@(t,y)f(t),tt,0,options);
tt = tt'; yy = yy';

% compute barycentric weights

ss = 2*mod(1:(N+1),2)-1;
ss(1)=.5*ss(1);
ss(N)=.5*ss(N);

% Interpolate using barycentric formula

t = linspace(-1,1,2000);
d = acos(t)*N/pi;
y = zeros(size(t));

for n=1:length(t)
    % check if t is one of the nodes
    k = round(d(n));
    if abs(d(n)-k) < 2*eps
        y(n) = yy(k+1);
    else
        r = ss./(t(n)-tt);
        y(n) = yy*r'/sum(r);
    end
end
```

```
figure
plot(t,y)
title('Polynomial of degree 10000')
```

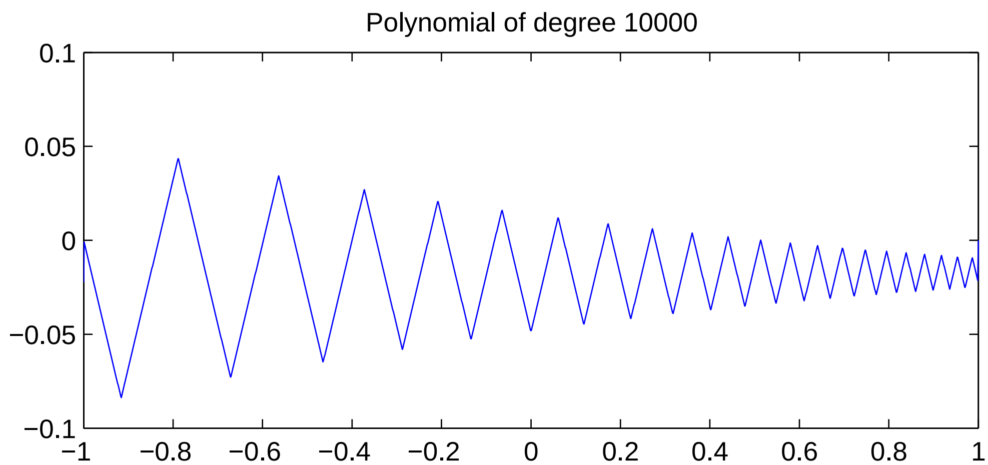

Figure 9: Chebyshev interpolant of degree 10000 computed on a grid of 2000 equispaced points. This example is taken from Trefethen's article "Six Myths of Polynomial Interpolation and Quadrature." (Myth #1)

This code interpolates the values of the function

$$F(t) = \int_0^t \operatorname{sign}\left(\sin\left(100\,\frac{s}{2-s}\right)\right)\,ds$$

at 10^4+1 Chebyshev nodes and computes the resulting polynomial of degree 10^4 at 2000 equispaced points all in about half a second. The interpolation itself takes very little time—about a tenth of a second!

The function is, obviously, not smooth and its interpolant does not look like a polynomial at all. Yet it is! Notice that there is no visible Runge phenomenon which would have swamped the computation on equispaced nodes. This code also shows the incredible power of the barycentric formula for Lagrangian interpolation. Let p be the Lagrangian polynomial with nodes

x_n, $n = 1, \ldots, N$. The barycentric formula for p is

$$p(x) = \frac{\displaystyle\sum_{n=1}^{N} f(x_n)\,\frac{w_n}{x - x_n}}{\displaystyle\sum_{n=1}^{N} \frac{w_n}{x - x_n}},$$

where w_n are the barycentric weights. For Chebyshev nodes the first and the last weights are $\frac{1}{2}$ and all other weights are 1—the simplicity of weights for Chebyshev nodes make them particularly attractive from the computational point of view.

Orthogonal polynomials

> We come now to the approximation process most commonly employed and most highly developed: least squares. An abstract vantage point from which it is convenient to survey the common features of various least square approximations is provided by the theory of inner product spaces. If the subjects of algebra, geometry, and analysis can be said to have a "center of gravity," it surely lies in this theory.
>
> *Philip J. Davis*

In addition to equioscillation, Chebyshev polynomials have many other remarkable properties. Chief among them is *orthogonality*. In order to understand what that means, we need to define inner product. This prompts another Linear Algebra interlude...

Inner product spaces

Chances are, you first saw vectors drawn as arrows. That is how vectors are typically introduced in Physics and Calculus. Later they acquire components and become n-tuples of numbers. In Multivariate Calculus the n-dimensional Euclidean space $\mathbb{R}^n$ is introduced essentially as a collection of arrows issuing from the origin. Each arrow has length—the usual expression given by the Pythagorian theorem. Furthermore, if two vectors are drawn, there is a well-defined (acute) angle between them. As one learns in Calculus, the easiest way to compute both the lengths and the angles in $\mathbb{R}^n$ is through the use of the *dot product*. The latter is commonly postulated by the formula

$$\langle x, y\rangle = x_1\, y_1 + x_2\, y_2 + \ldots + x_n\, y_n, \tag{2}$$

which we will write in an equivalent form:

$$\langle x, y\rangle = x^T\, y.$$

Recall that the superscript "T" is for 'transpose', which is to say, "interchange rows with columns"; it is the single quote in `Matlab`:

```
s = x'*y; % dot product of two vectors
```

With the dot product defined by (2), it is easy to see that the length of a vector should be:

$$|x| = \sqrt{\langle x, x\rangle}.$$

Meanwhile, the angle between x and y can be computed from the relation:

$$\cos\theta = \frac{\langle x, y\rangle}{|x|\,|y|}.$$

The latter formula is not obvious; it is usually justified using the Law of Cosines from geometry.

Have you ever wondered why the dot product is defined by Equation (2)? It does not have to be. The dot product on $\mathbb{R}^n$ is just a special case of an *inner product*. In general, an inner product $\langle\cdot,\cdot\rangle$ is a mapping which takes a pair of vectors into a number subject to the following axioms:

(IP1) Inner product of any vector with itself must be a non-negative number: $\langle x, x\rangle \geq 0$. Moreover, $\langle x, x\rangle = 0$ implies $x = 0$.

(IP2) Inner product is symmetric: $\langle x, y\rangle = \langle y, x\rangle$.

(IP3) Inner product is linear in the first argument: $\langle \sum_n c_n\, x_n, y\rangle = \sum_n c_n \langle x_n, y\rangle$; as a consequence of symmetry, inner product is linear in the second argument as well.

Any bilinear mapping that takes a pair of vectors into a scalar is called a bilinear *form*. One can summarize the conditions imposed on an inner product in one sentence as follows: Inner product is a positive definite, symmetric, bilinear form.

As was the case with norms, the concept of inner product makes sense only within some vector space. A vector space equipped with an inner product is called, naturally, an *inner product space*. And, speaking of norms, an inner product space has a natural norm coming from the inner product—the 2-norm:

$$\|x\| = \sqrt{\langle x, x\rangle}.$$

Thus any inner product space is automatically a normed space. The converse is not true, however: most norms are not derived from inner products.

As you may have already guessed, the (acute) angle between two vectors in an inner product space is defined by the equation

$$\cos(\theta) = \frac{\langle x, y\rangle}{\sqrt{\langle x, x\rangle}\,\sqrt{\langle y, y\rangle}}. \tag{3}$$

This is the same formula as in Multivariate Calculus but with new angle notation. If $\langle x, y\rangle = 0$ then, clearly, $\cos(\theta) = 0$ and the vectors are perpendicular. We will commonly use the term 'orthogonal' instead of 'perpendicular'.

We conclude this section with a justification of the formula (3). At this point it may not be clear why the angle between vectors may be defined in that manner. Indeed, if we have some strange inner product, how do we know that the right-hand side of (3) is always in $[-1, 1]$? For justification, take two arbitrary vectors x and y and consider the inner product

$$\langle x - t\,y, x - t\,y\rangle$$

as a function of the scalar t. Using bilinearity and symmetry, we can rewrite this inner product as the quadratic:

$$q(t) = \langle x, x\rangle - 2\,\langle x, y\rangle\, t + \langle y, y\rangle\, t^2.$$

On the other hand, by positive-definitiveness $q(t)$ has imaginary roots. Therefore the discriminant is non-positive:

$$4\,\langle x, y\rangle^2 - 4\,\langle x, x\rangle\,\langle y, y\rangle \le 0.$$

Yet that means

$$\frac{\langle x, y\rangle^2}{\langle x, x\rangle\,\langle y, y\rangle} \le 1$$

and consequently Equation (3) is well-defined.

Examples of inner products

Whenever we work with vectors in $\mathbb{R}^n$ we usually use the dot product (2) as the default inner product. However, on occasion, we may prefer to use other inner products. For instance, the following is an inner product on $\mathbb{R}^3$ which is clearly different from the dot product:

$$\langle x, y \rangle = x_1\, y_1 + 2\, x_2\, y_2 + 3\, x_3\, y_3. \tag{4}$$

We can call it a *weighted dot product*. As an easy exercise, verify that Equation (4) satisfies all of the necessary axioms.

As a more pertinent example to Numerical Analysis, let V be the space of functions of one variable defined on the interval $[-1, 1]$. Define the inner product on V by:

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\, g(x)\, dx \tag{5}$$

We will call the inner product (5) the L^2-inner product (L is a tribute to Lebesgue who developed the theory of integration); it is the continuous analogue of the dot product (2).

Finally, a weighted L^2-inner product on $[-1, 1]$ is defined by

$$\langle f, g \rangle = \int_{-1}^{1} f(x)\, g(x)\, w(x)\, dx \tag{6}$$

In (6) the weight $w(x)$ must be nonnegative with

$$0 < \int_{-1}^{1} w(x)\, dx, \infty.$$

As we will soon show, Chebyshev polynomials are orthogonal on $[-1, 1]$ with respect to a certain weighted L^2-inner product. First, however, let us discuss orthogonal expansions.

Orthogonal expansions

Let V be an n-dimensional vector space equipped with a basis $\mathcal{B} = \{v_1, \ldots, v_n\}$. Consider the problem of finding the components of $w \in V$ with respect the the basis $\mathcal{B}$. That is, we want to find coefficients in the linear combination:

$$w = c_1\, v_1 + \ldots + c_n\, v_n. \tag{7}$$

Generally, this can be a very difficult problem. However, let us equip V with an inner product $\langle \cdot, \cdot \rangle$—any inner product. Taking the inner products of both sides of Equation (7) with the basis vectors, we get n relations of the form

$$\langle w, v_m \rangle = c_1 \langle v_1, v_m \rangle + \ldots + c_n \langle v_n, v_m \rangle. \tag{8}$$

These are the usual linear equations with scalar coefficients and scalar unknowns which we can write in matrix form as:

$$G\, c = b.$$

Here the n-by-n matrix G has entries $\langle v_n, v_m \rangle$ and is called Gram's matrix; the right-hand side vector b has known components $\langle w, v_m \rangle$ and c is the vector of coefficients that are to be determined. The standard linear system $G\, c = b$ can be solved through Gaussian elimination; symbolically, $c = G^{-1} b$.

Gram's matrices $G = \langle v_n, v_m \rangle$ have a number of interesting properties. They are clearly *symmetric*: $G^T = G$. More importantly, they are nonsingular as long as the vectors $\{v_n\}$ are linearly independent, which must be true if $\{v_n\}$ form a basis. Some of the exercises at the end ask you to confirm these and other properties of symmetric matrices computationally.

Having dealt with the general problem of basis expansion in inner product spaces, let us consider a particularly important special case where the basis vectors are mutually orthogonal: $\langle v_n, v_m \rangle = 0$ if $n \neq m$. It is easy to see that the Gram matrix in this case is diagonal with diagonal entries $\langle v_n, v_n \rangle$. Consequently, the explicit solution of the expansion problem in this case is:

$$w = \frac{\langle w, v_1 \rangle}{\langle v_1, v_1 \rangle}\, v_1 + \ldots + \frac{\langle w, v_n \rangle}{\langle v_n, v_n \rangle}\, v_n \tag{9}$$

Equation (10) is called "orthogonal basis expansion" or "Fourier expansion". It further simplifies to

$$w = \langle w, v_1 \rangle\, v_1 + \ldots + \langle w, v_n \rangle\, v_n$$

if the basis vectors have unit length. Henceforth we will call bases of orthogonal vectors simply 'orthogonal bases' and the corresponding basis expansions 'orthogonal expansions'. Orthogonal bases where vectors have unit length are usually called 'orthonormal.'

As a simple example, consider the standard basis of $\mathbb{R}^3$: $\{i, j, k\}$. It is an orthonormal basis with respect to the standard dot product. Therefore, any vector $w \in \mathbb{R}^3$ can be written as:

$$w = \langle w, i \rangle\, i + \langle w, j \rangle\, j + \langle w, k \rangle\, k.$$

As another example, consider quadratics on $[-1,1]$ with L^2-inner product (5). The *Legendre polynomials* $\{P_0 = 1, P_1 = x, P_2 = \frac{3}{2}x^2 - \frac{1}{2}\}$ are easily seen to be orthogonal (but not orthonormal). Therefore, any quadratic q can be written as the following combination:

$$q = \frac{\langle q, P_0\rangle}{\langle P_0, P_0\rangle} P_0 + \frac{\langle q, P_1\rangle}{\langle P_1, P_1\rangle} P_1 + \frac{\langle q, P_2\rangle}{\langle P_2, P_2\rangle} P_2. \tag{10}$$

For instance, let $q = x^2 + x + 1$. Then

$$\begin{aligned}
\langle q, P_0\rangle &= \int_{-1}^{1} (x^2 + x + 1)\, 1\, dx = \frac{8}{3} \\
\langle q, P_1\rangle &= \int_{-1}^{1} (x^2 + x + 1)\, x\, dx = \frac{2}{3} \\
\langle q, P_2\rangle &= \int_{-1}^{1} (x^2 + x + 1) \left(\frac{3}{2}x^2 - \frac{1}{2}\right) dx = \frac{4}{15}
\end{aligned}$$

Furthermore,

$$\begin{aligned}
\langle P_0, P_0\rangle &= \int_{-1}^{1} 1^2\, dx = 2 \\
\langle P_1, P_1\rangle &= \int_{-1}^{1} x^2\, dx = \frac{2}{3} \\
\langle P_2, P_2\rangle &= \int_{-1}^{1} \left(\frac{3}{2}x^2 - \frac{1}{2}\right)^2 dx = \frac{2}{5}
\end{aligned}$$

Finally,

$$\begin{aligned}
\frac{\langle q, P_0\rangle}{\langle P_0, P_0\rangle} P_0 + \frac{\langle q, P_1\rangle}{\langle P_1, P_1\rangle} P_1 + \frac{\langle q, P_2\rangle}{\langle P_2, P_2\rangle} P_2 &= \frac{(8/3)}{2}\, 1 + \frac{(2/3)}{(2/3)}\, x + \frac{(4/15)}{2/5} \left(\frac{3}{2}x^2 - \frac{1}{2}\right) \\
&= \frac{4}{3} + x + x^2 - \frac{1}{3} = q,
\end{aligned}$$

as required.

Fourier approximation

We have shown that any quadratic can be written as an orthogonal expansion of the first three Legendre polynomials:

$$q = \frac{\langle q, P_0 \rangle}{\langle P_0, P_0 \rangle} P_0 + \frac{\langle q, P_1 \rangle}{\langle P_1, P_1 \rangle} P_1 + \frac{\langle q, P_2 \rangle}{\langle P_2, P_2 \rangle} P_2. \qquad (11)$$

Suppose that q is not a quadratic but is some function on $[-1, 1]$. Although Equation (11) clearly cannot hold in that case, its right-hand side still makes sense and can be regarded as a quadratic approximation to q. For example, let $q = \cos(x)$. Then, inspired by Equation (11), we can approximate the cosine on $[-1, 1]$ as follows:

$$\cos(x) \approx \sin(1) + \frac{5}{2}\,(-4\,\sin(1) + 6\,\cos(1))\left(\frac{3}{2}x^2 - \frac{1}{2}\right)$$

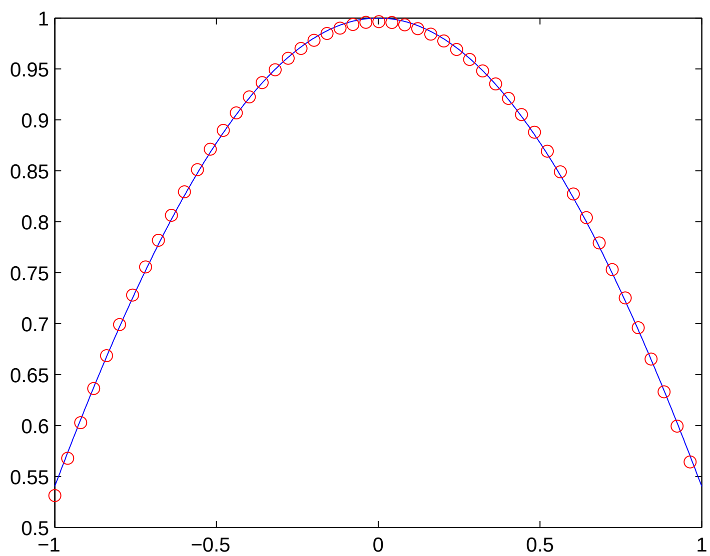

Figure 10: Approximation of $\cos(x)$ with a linear combination of the first three Legendre polynomials

As Figure 10 clearly suggests, the approximation is excellent throughout the interval. Compare that with Figure 11 below which shows the Taylor approximation.

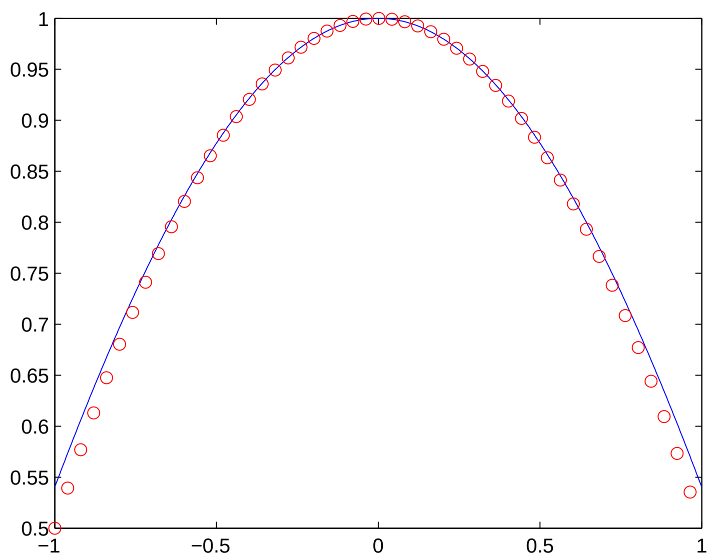

Figure 11: Approximation of $\cos(x)$ with the Taylor quadratic

The Taylor approximation is slightly better near the origin—the center of expansion. However, the accuracy of Taylor approximation quickly degrades away from the origin. In contrast, the accuracy of the Fourier approximation is consistent throughout the interval.

As another classic example, consider the space of functions on $[0, \pi]$ which vanish at the endpoints and which have "integrable squares":

$$V = \{f(x) \mid \int_0^\pi f(x)^2 \, dx < \infty\}$$

The integrability of squares allows for the L^2 inner product on V:

$$\langle f, g \rangle = \int_0^\pi f(x) \, g(x) \, dx.$$

You can easily check (exercise) that the functions

$$\phi_n(x) = \sqrt{\frac{2}{\pi}}\,\sin(n\,x), \quad n = 1, 2, 3, \ldots$$

belong to V, are orthogonal under the L^2-inner product, and have unit L^2-norm. Consequently, any function in V can be approximated with a sum of sines of integer frequencies:

$$f(x) \approx \sum_{n=1}^{N} \langle f, \phi_n \rangle\, \phi_n(x).$$

The greater N is, the better is the approximation. Furthermore, the approximation can be used for functions which do not vanish at the endpoints. As an example, take $f(x) = x$, which is clearly not a combination of sines.

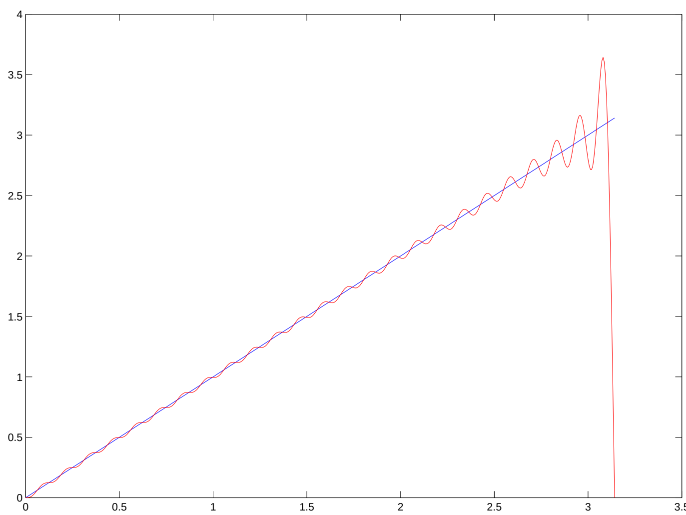

Figure 12: Approximation of x on $[0, \pi]$ with sines of integer frequencies ($N = 50$); the oscillations near $x = \pi$ are due to the fact that all sines vanish there.

Using integration by parts, one can show that

$$\langle f, \phi_n \rangle = -\sqrt{\frac{2}{\pi}}\,\frac{\pi}{n}\,\cos(\pi\,n)$$

Consequently, after simplification, we can write:

$$x \approx \sum_{n=1}^{N} \left(-\frac{2\,\cos(\pi\,n)}{n} \right)\,\sin(n\,x) \tag{12}$$

Figure 12 shows the Fourier approximation (12) with $N = 50$. The Fourier sum struggles to converge near $x = \pi$ where all sines vanish but the function has value π. This is an example of Gibbs' phenomenon—a close cousin of Runge's phenomenon.

Orthogonality of Chebyshev polynomials

Chebyshev polynomials of the first kind T_n are orthogonal on $[-1, 1]$ with respect to the weighted L^2-inner product with weight: $w(x) = (1 - x^2)^{-1/2}$. That is:

$$\langle T_n, T_m \rangle = \int_{-1}^{+1} T_n(x)\,T_m(x)\frac{dx}{\sqrt{1-x^2}} = 0, \quad \text{if } n \neq m.$$

To see that, substitute $x = \cos(t)$. This leads to the integral of the product of two cosines

$$\int_0^{\pi} \cos(n\,t)\,\cos(m\,t)\,dt$$

which can be evaluated using the addition formula for cosines (exercise). Consequently, any reasonable function on $[-1, 1]$ can be approximated with a sum of Chebyshev polynomials:

$$f(x) \approx \sum_{n=0}^{N} \frac{\langle f, T_n \rangle}{\langle T_n, T_n \rangle}\,T_n(x).$$

If we set $x = \cos(t)$ then

$$f(\cos(t)) \approx \sum_{n=0}^{N} \frac{\langle f, T_n \rangle}{\langle T_n, T_n \rangle}\,\cos(n\,t).$$

Thus Chebyshev expansion of $f(x)$ is really the same as Fourier cosine expansion of $f(\cos(t))$.

Gram-Schmidt orthogonalization

Let V be an inner product space. Suppose that $\{v_1, \ldots, v_n\}$ are linearly independent, but not necessarily orthogonal vectors in V. Gram-Schmidt orthogonalization is a process that converts these vectors into *orthonormal* vectors $\{w_1, \ldots, w_n\}$. The first step is to normalize the first vector:

$$w_1 = \frac{v_1}{\|v_1\|}.$$

Of course, the norm comes from the inner product: $\|v_1\| = \sqrt{\langle v_1, v_1 \rangle}$. Next one computes the vector

$$z_2 = v_2 - \langle v_2, w_1 \rangle\, w_1.$$

Geometrically, we subtract from v_2 its orthogonal projection onto w_1. The remainder is a vector orthogonal to w_1. Indeed:

$$\langle w_1, z_2 \rangle = \langle w_1, v_2 \rangle - \langle v_2, w_1 \rangle \langle w_1, w_1 \rangle = 0, \quad (\langle w_1, w_1 \rangle = 1).$$

We now have two orthogonal vectors but the norm of z_2 is not necessarily one. This is easy to correct, however. Define:

$$w_2 = \frac{z_2}{\|z_2\|}$$

In the same manner, to compute w_3 one first computes:

$$z_3 = v_3 - \langle v_3, w_1 \rangle\, w_1 - \langle v_3, w_2 \rangle\, w_2.$$

This gives a vector orthogonal to w_1 and w_2 (exercise: check this!) Next normalize z_3:

$$w_3 = \frac{z_3}{\|z_3\|}.$$

In general, having computed $w_1, \ldots, w_k$, one adds the next vector via:

$$\begin{aligned} z_{k+1} &= v_{k+1} - \sum_{j=1}^{k} \langle v_{k+1}, w_j \rangle\, w_j \\ w_{k+1} &= \frac{z_{k+1}}{\|z_{k+1}\|} \end{aligned}$$

The assumption of linear independence ensures that there is no division by zero.

Legendre polynomials

Legendre polynomials are orthogonal on $[-1, 1]$ with respect to the standard L^2-inner product. They can be generated by orthogonalizing the monomials $\{1, x, x^2, x^3 \ldots\}$. The `Maple` code below generates the first five normalized Legendre polynomials:

```
> L2InnerProduct := (f,g) -> int(f*g,x=-1..1);
> L2Norm := f -> sqrt(L2InnerProduct(f,f));
> V := [seq(x^(k-1),k=1..5)];
> W[1] := V[1]/L2Norm(V[1]);
> for n from 2 to 5 do
>     z := V[n] - add(L2InnerProduct(V[n],W[j])*W[j],j=1..n-1);
>     W[n] := z/L2Norm(z);
>     print(W[n]);
> end do:
```

The output is:

$$\begin{aligned}
W_1 &= 1/2\,\sqrt{2} \\
W_2 &= 1/2\,\sqrt{6}x \\
W_3 &= 3/4\,\left(x^2 - 1/3\right)\sqrt{10} \\
W_4 &= 5/4\,\left(x^3 - 3/5\,x\right)\sqrt{14} \\
W_5 &= \frac{105}{16}\left(x^4 + \frac{3}{35} - 6/7\,x^2\right)\sqrt{2}
\end{aligned}$$

Classical Legendre polynomials P_n have a different normalization and do not have unit length. The first five classical Legendre polynomials are:

$$1, \quad x, \quad \frac{3}{2}x^2 - \frac{1}{2}, \quad \frac{5}{2}x^3 - \frac{3}{2}x, \quad \frac{35}{8}x^4 - \frac{15}{4}x^2 + \frac{3}{8}$$

These can be generated in `Maple` using:

```
> with(orthopoly);
> for n from 0 to 4 do
>     P(n,x);
> end do;
```

Exercises

1. Think of vectors x and y in $\mathbb{R}^2$ as, simply, points in the plane. Given a norm $\|\cdot\|$ on $\mathbb{R}^2$, we can define distance from x to y as $\|x-y\|$. A circle of radius r centered at the origin can then be defined as a set of points equidistant from the origin, which is to say all vectors of norm one. Plot the unit circles corresponding to the p-norms with $p = 1, 2, 4, \infty$.

2. Define the inner product on $\mathbb{R}^2$ by the formula:

$$\langle x, y \rangle = x_1\, y_1 + 4\, x_2\, y_2.$$

   Here we use the convention of denoting the n-th component of a vector with a subscript. Find a two-by-two matrix A such that:

$$\langle x, y \rangle = x^T\, A\, y.$$

3. Let

$$A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}$$

   For $x, y \in \mathbb{R}^2$ suppose we define

$$[x, y] = x^T\, A\, y.$$

   Is $[x, y]$ an inner product on $\mathbb{R}^2$? If you answered "No", explain what breaks down. If you answered "Yes" either provide a formal argument, if you can, or generate convincing evidence in `Matlab`.

4. Consider quadratics on the interval $[0, 1]$ with the L^2 inner product:

$$\langle f, g \rangle = \int_0^1 f(x)\, g(x)\, dx.$$

   Find an orthogonal basis of this space. Validate your result with an orthogonal expansion of $q = x^2 + x + 1$.

5. Consider Lagrange interpolation on the interval $[-1, 1]$. Suppose we want to minimize the remainder by choosing nodes so as to minimize the norm of the nodal polynomial p. From the handout you know that minimization of the ∞-norm of p leads to the conclusion that p must be a (monic) Chebyshev polynomial of the first kind. Show that minimization of the 2-norm leads to (monic) Legendre polynomials.

6. Which monic polynomials have the smallest 1-norm on $[-1,1]$? Compute the first three numerically and try to identify them using `Maple`. 50 bonus points for a rigorous proof.

7. Define a weighted L^2-inner product on $[-1,1]$ as

$$\langle f, g\rangle = \int_{-1}^{1} f(x)\, g(x)\, (1-x^2)\, dx.$$

Find the first ten orthogonal polynomials corresponding to that inner product. Let p_9 be the last of these—polynomial of degree nine. Find its roots and use them as nodes to interpolate

$$y = \frac{1}{1+25\,x^2}.$$

Compare the result with interpolation of the same function on equispaced nodes. What are your observations?