Chapter 5

Inner Product Spaces

Up to this point all the vectors that we have looked at have been vectors in $\mathbb{R}^n$, but more abstractly a vector can be any object in a set that satisfies the axioms of a vector space. We will not go into the rigorous details of what constitutes a vector space, but the essential idea is that a vector space is any set of objects where it is possible to form linear combinations of those objects[1] in a reasonable way and get another object in the same set as a result. In this chapter we will be looking at examples of some vector spaces other than $\mathbb{R}^n$ and at generalizations of the dot product on these spaces.

5.1 Inner Products

Definition 16 *Given a vector space, V, an* **inner product** *on V is a rule (which must satisfy the conditions given below) for multiplying elements of V together so that the result is a scalar. If $\mathbf{u}$ and $\mathbf{v}$ are vectors in V, then their inner product is written $\langle \mathbf{u}, \mathbf{v} \rangle$. The inner product must satisfy the following conditions for any $\mathbf{u}$, $\mathbf{v}$, $\mathbf{w}$ in V and any scalar c.*

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$
4. $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ *and* $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ *if and only if* $\mathbf{u} = \mathbf{0}$.

A vector space with an inner product is called an **inner product space**.

You should realize that these four conditions are satisfied by the dot product (or standard inner product on $\mathbb{R}^n$), but as we will soon see there are many other examples of inner products.

Example 5.1.1

Let $\mathbf{u} = \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} v_1 \\ v_2 \end{bmatrix}$ be vectors in $\mathbb{R}^2$. Define

$$\langle \mathbf{u}, \mathbf{v} \rangle = 2u_1v_1 + 4u_2v_2$$

This rule defines an inner product on $\mathbb{R}^2$. To verify this it is necessary to confirm that all 4 conditions given in the definition of an inner product are satisfied. We will just look at conditions 2 and 4 and leave the other two as an exercise.

[1] That is, addition and scalar multiplication are defined on these objects.

For condition 2 we have

$$\begin{aligned}\langle \mathbf{u}+\mathbf{v},\mathbf{w}\rangle &= 2(u_1+v_1)w_1+4(u_2+v_2)w_2\\ &= 2u_1w_1+2v_1w_1+4u_2w_2+4v_2w_2\\ &= (2u_1w_1+4u_2w_2)+(2v_1w_1+4v_2w_2)\\ &= \langle \mathbf{u},\mathbf{w}\rangle+\langle \mathbf{v},\mathbf{w}\rangle\end{aligned}$$

For condition 4 we have

$$\langle \mathbf{u},\mathbf{u}\rangle = 2u_1^2+4u_2^2 \geq 0$$

since the sum of two squares cannot be negative.
Furthermore, $2u_1^2+4u_2^2=0$ if and only if $u_1=0$ and $u_2=0$. That is, $\langle \mathbf{u},\mathbf{u}\rangle=0$ if and only if $\mathbf{u}=\mathbf{0}$.

If you look at the last example you should see a similarity between the inner product used there and the dot product. The dot product of $\mathbf{u}$ and $\mathbf{v}$ would be $u_1v_1+u_2v_2$. The example given above still combines the same two terms but now the terms are weighted. As long as these weights are positive numbers this procedure will always produce an inner product in $\mathbb{R}^n$ by a simple modification of the dot product. This type of inner product is called a **weighted dot product**.

This variation on the dot product can be written another way. The dot product of $\mathbf{u}$ and $\mathbf{v}$ can be written $\mathbf{u}^T\mathbf{v}$. A weighted dot product can be written $\mathbf{u}^TD\mathbf{v}$ where D is a diagonal matrix with positive entries on the diagonal. (Which of the four conditions of an inner product would not be satisfied if the weights were not positive?)

The next example illustrates an inner product on a vector space other than $\mathbb{R}^n$.

Example 5.1.2

The vector space P_n is the vector space of polynomials of degree less than or equal to n. In particular, P_2 is the vector space of polynomials of degree less than or equal to 2. If p and q are vectors in P_2 define

$$\langle p,q\rangle = p(-1)q(-1)+p(0)q(0)+p(1)q(1)$$

This rule will define an inner product on P_2. It can be verified that all four conditions of an inner product are satisfied but we will only look at conditions 1 and 4, leaving the others as an exercise.
For condition 1:

$$\begin{aligned}\langle p,q\rangle &= p(-1)q(-1)+p(0)q(0)+p(1)q(1)\\ &= q(-1)p(-1)+q(0)p(0)+q(1)p(1)\\ &= \langle q,p\rangle\end{aligned}$$

For condition 4:

$$\langle p,p\rangle = [p(-1)]^2+[p(0)]^2+[p(1)]^2$$

It is clear that this expression, being the sum of three squares, is always greater than or equal to 0, so we have $\langle p,p\rangle \geq 0$.
Next we want to show that $\langle p,p\rangle=0$ if and only if $p=0$. It's easy to see that if $p=0$ then $\langle p,p\rangle=0$.

On the other hand, suppose $\langle p,p\rangle=0$, then we must have $p(-1)=0$, $p(0)=0$, and $p(1)=0$. This means p has 3 roots, but since p has degree less than or equal to 2 the only way this is possible is if $p=0$, that is, p is the zero polynomial.

Suppose we had $p(t) = 2t^2 - t + 1$ and $q(t) = 2t - 1$, then

$$\langle p, q\rangle = (4)(-3) + (1)(-1) + (2)(1) = -11$$

Also

$$\langle p, p\rangle = 4^2 + 1^2 + 2^2 = 21$$

Evaluating an inner product of two polynomial using this rule can be broken down into two steps:

- **Step 1**: You should first sample the polynomials at the values -1, 0, and 1. The samples of each polynomial will give you a vector in $\mathbb{R}^3$.
- **Step 2**: You take the dot product of the two vectors created by sampling. (As a variation, this step could be a weighted inner product).

The significance of inner product spaces is that when you have a vector space with an inner product then all of the ideas covered in the previous chapter connected with the dot product (such as length, distance, and orthogonality) can now be applied to the inner product space. In particular, the Orthogonal Decomposition Theorem and the Best Approximation Theorem are true in any inner product space (where any expression involving a dot product is replaced by an inner product).

We will now list some basic definitions that can be applied to any inner product space.

- We define the *length* or *norm* of a vector $\mathbf{v}$ in an inner product space to be

$$\|\mathbf{v}\| = \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle}$$

- A *unit vector* is a vector of length 1.
- The *distance* between $\mathbf{u}$ and $\mathbf{v}$ is defined as $\|\mathbf{u} - \mathbf{v}\|$.
- The vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.
- The orthogonal projection of $\mathbf{u}$ onto a subspace W with orthogonal basis $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k\}$ is given by $\text{Proj}_W \mathbf{u} = \displaystyle\sum_{i=1}^{k} \frac{\langle \mathbf{u}, \mathbf{v}_i\rangle}{\langle \mathbf{v}_i, \mathbf{v}_i\rangle}\mathbf{v}_i$.

We will next prove two fundamental theorems which apply to any inner product space.

Theorem 5.1 (The Cauchy-Schwarz Inequality) *For all vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *in an inner product space* V *we have*

$$|\langle \mathbf{u}, \mathbf{v}\rangle| \leq \|\mathbf{u}\|\|\mathbf{v}\|$$

Proof. There are various ways of proving this theorem. The proof we give is not the shortest but it is straightforward. Each of the four conditions of an inner product must be used in the proof. You should try to justify each step of the proof and discover where each of thse rules are used.

If either $\mathbf{u} = \mathbf{0}$ or $\mathbf{v} = \mathbf{0}$ then both sides of the given inequality would be zero and the inequality would therefore be true. (Here we are basically saying that $\langle \mathbf{0}, \mathbf{v}\rangle = 0$. This is not as trivial as it might seem. Which condition of an inner product jusifies this statement?)

We will now assume that both $\mathbf{u}$ and $\mathbf{v}$ are non-zero vectors. The inequality that we are trying to prove can be written as

$$-\|\mathbf{u}\|\,\|\mathbf{v}\| \leq \langle \mathbf{u}, \mathbf{v}\rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$$

We proceed as follows (you should try to find the justification for each of the following steps): First we can say

$$\langle \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|} \rangle \geq 0$$

We also have

$$\begin{aligned} \langle \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{u}}{\|\mathbf{u}\|} - \frac{\mathbf{v}}{\|\mathbf{v}\|} \rangle &= \frac{\langle \mathbf{u}, \mathbf{u} \rangle}{\|\mathbf{u}\|^2} - 2\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} + \frac{\langle \mathbf{v}, \mathbf{v} \rangle}{\|\mathbf{v}\|^2} \\ &= \frac{\|\mathbf{u}\|^2}{\|\mathbf{u}\|^2} - 2\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} + \frac{\|\mathbf{v}\|^2}{\|\mathbf{v}\|^2} \\ &= 2 - 2\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \end{aligned}$$

If we put the last two results together we get

$$2 - 2\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \geq 0$$

Rearranging this last inequality we have

$$2\frac{\langle \mathbf{u}, \mathbf{v} \rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \leq 2$$

and therefore

$$\langle \mathbf{u}, \mathbf{v} \rangle \leq \|\mathbf{u}\|\,\|\mathbf{v}\|$$

The proof is not finished. We still have to show that

$$\langle \mathbf{u}, \mathbf{v} \rangle \geq -\|\mathbf{u}\|\,\|\mathbf{v}\|$$

We will leave it to you to fill in the details but the remaining part of the proof is a matter of repeating the above argument with the first expression replaced with

$$\langle \frac{\mathbf{u}}{\|\mathbf{u}\|} + \frac{\mathbf{v}}{\|\mathbf{v}\|}, \frac{\mathbf{u}}{\|\mathbf{u}\|} + \frac{\mathbf{v}}{\|\mathbf{v}\|} \rangle$$

■

Theorem 5.2 (The Triangle Inequality) *For all vectors* $\mathbf{u}$ *and* $\mathbf{v}$ *in an inner product space* V *we have*

$$\|\mathbf{u} + \mathbf{v}\| \leq \|\mathbf{u}\| + \|\mathbf{v}\|$$

Proof. The following lines show the basic steps of the proof. We leave it to the reader to fill in the justifications of each step (the Cauchy-Schwarz inequality is used at one point).

$$\begin{aligned} \|\mathbf{u} + \mathbf{v}\|^2 &= \langle \mathbf{u} + \mathbf{v}, \mathbf{u} + \mathbf{v} \rangle \\ &= \langle \mathbf{u}, \mathbf{u} \rangle + 2\langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{v}, \mathbf{v} \rangle \\ &\leq \|\mathbf{u}\|^2 + 2|\langle \mathbf{u}, \mathbf{v} \rangle| + \|\mathbf{v}\|^2 \\ &\leq \|\mathbf{u}\|^2 + 2\|\mathbf{u}\|\,\|\mathbf{v}\| + \|\mathbf{v}\|^2 \\ &= (\|\mathbf{u}\| + \|\mathbf{v}\|)^2 \end{aligned}$$

We now take the square root of both sides and the inequality follows. ■

One important consequence of the Cauchy-Schwarz inequality is that $-1 \le \dfrac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|} \le 1$ for non-zero vectors **u** and **v**. This makes it reasonable to define the angle between two non-zero vectors **u** and **v** in an inner product space as the unique value of θ such that

$$\cos\theta = \frac{\langle \mathbf{u}, \mathbf{v}\rangle}{\|\mathbf{u}\|\,\|\mathbf{v}\|}$$

Example 5.1.3

In P_2 with the inner product defined by

$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

Let $p(t) = t^2 + t + 1$. Find a unit vector orthogonal to p.

The vector that we are looking for must have the form $q(t) = at^2 + bt + c$ for some scalars a, b, c. Since we want q to be orthogonal to p we must have $\langle p, q\rangle = 0$. This results in

$$\begin{aligned} p(-1)q(-1) + p(0)q(0) + p(1)q(1) &= \\ (1)(a - b + c) + (1)(c) + (3)(a + b + c) &= \\ 4a + 2b + 5c &= 0 \end{aligned}$$

We can use any values of a, b, and c which satisfy this last condition. For example, we can use $a = 2$, $b = 1$, and $c = -2$ giving $q(t) = 2t^2 + t - 2$. But this is not a unit vector, so we have to normalize it. We have

$$\langle q, q\rangle = (-1)^2 + (-2)^2 + 1^2 = 6$$

We now conclude that $\|q\| = \sqrt{6}$, so by normalizing q we get the following unit vector orthogonal to p:

$$\frac{1}{\sqrt{6}}\left(2t^2 + t - 2\right)$$

We are dealing here with abstract vector spaces and although you can transfer some of your intuition from $\mathbb{R}^n$ to these abstract spaces you have to be careful. In this example there is nothing in the graphs of $p(t)$ and $q(t)$ that reflects their orthogonality relative to the given inner product.

Example 5.1.4

In P_3 the rule

$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

will not define an inner product. To see this let $p(t) = t^3 - t$, then the above rule would give $\langle p, p\rangle = 0^2 + 0^2 + 0^2 = 0$ which contradicts condition 4. (Basically this is because a cubic, unlike a quadratic, *can* have roots at -1, 0 and 1.)
On the other hand if we modify the formula slightly we can get an inner product on P_3. We just let

$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1) + p(2)q(2)$$

We leave the confirmation that this defines an inner product to the reader, but we will mention a few points connected with this inner product:

- When a polynomial is sampled at n points you can look at the result as a vector in $\mathbb{R}^n$, the given inner product is then equivalent to the dot product of these vectors in $\mathbb{R}^n$.
- The points where the functions are being sampled are unimportant in a sense. Instead of sampling at -1, 0, 1, and 2 we could have sampled at 3, 5, 6, and 120 and the result would still be an inner product. The actual value of the inner product of two specific vectors would vary depending on the sample points.
- To define an inner product in this way you need to sample the polynomial at more points than the highest degree allowed. So, for example, in P_5 you would have to sample the polynomials at at least 6 points.

Example 5.1.5

I n P_3 define an inner product by sampling at $-1, 0, 1, 2$. Let

$$p_1(t) = t - 3\ ,\ p_2(t) = t^2 - 1\ ,\ q(t) = t^3 - t^2 - 2$$

Sampling these polynomials at the given values would give the following:

$$p_1 \to \begin{bmatrix} -4 \\ -3 \\ -2 \\ -1 \end{bmatrix} \qquad p_2 \to \begin{bmatrix} 0 \\ -1 \\ 0 \\ 3 \end{bmatrix} \qquad q \to \begin{bmatrix} -4 \\ -2 \\ -2 \\ 2 \end{bmatrix}$$

First notice that p_1 and p_2 are orthogonal since

$$\langle p_1, p_2 \rangle = (-4)(0) + (-3)(-1) + (-2)(0) + (-1)(3) = 0$$

Now let $W = \text{Span}\{p_1, p_2\}$. We will find $\text{Proj}_W q$. Since we have an orthogonal basis of W we can compute this projection as

$$\frac{\langle q, p_1 \rangle}{\langle p_1, p_1 \rangle} p_1 + \frac{\langle q, p_2 \rangle}{\langle p_2, p_2 \rangle} p_2$$

This gives

$$\frac{24}{30}(t - 3) + \frac{8}{10}(t^2 - 1) = \frac{4}{5}t^2 + \frac{4}{5}t - \frac{16}{5}$$

The orthogonal component of this projection would then be

$$(t^3 - t^2 - 2) - (\frac{4}{5}t^2 + \frac{4}{5}t - \frac{16}{5}) = t^3 - \frac{9}{5}t^2 - \frac{4}{5}t + \frac{6}{5}$$

You should confirm for yourself that this last result is orthogonal to both p_1 and p_2 as expected.

One of the most important inner product spaces is the vector space $C\,[a, b]$ of continuous functions on the interval $a \leq t \leq b$ with an inner product defined as

$$\langle f, g \rangle = \int_a^b f(t)g(t)\,dt$$

The first three conditions of an inner product follow directly from elementary properties of definite integrals. For condition 4 notice that

$$\langle f, f\rangle = \int_a^b [f(t)]^2 \, dt \geq 0$$

The function $[f(t)]^2$ is continuous and non-negative on the interval from a to b. The details of verifying condition 4 would require advanced calculus, but the basic idea is that if the integral over this interval is 0, then the area under the curve must be 0, and so the function itself must be identically 0 since the function being integrated is never negative.

Example 5.1.6

In $C[0, \pi/2]$ with the inner product

$$\langle f, g\rangle = \int_0^{\pi/2} f(t)g(t)\, dt$$

Let $f(t) = \cos t$ and $g(t) = \sin t$. Find the projection of f onto g.
The point here is that you would follow the same procedure for finding the projection of one vector onto another that you already know except that the dot product gets replaced by the inner product. We will represent the projection as $\hat{f}$. We then have

$$\begin{aligned}
\hat{f} &= \frac{\langle f, g\rangle}{\langle g, g\rangle} g \\
&= \frac{\int_0^{\pi/2} \cos t \sin t \, dt}{\int_0^{\pi/2} \sin^2 t \, dt} \sin t \\
&= \frac{1/2}{\pi/4} \sin t \\
&= \frac{2}{\pi} \sin t
\end{aligned}$$

The orthogonal component of the projection would be

$$f - \hat{f} = \cos t - \frac{2}{\pi} \sin t$$

Example 5.1.7

In $C[0, 1]$ let $f_1(t) = t^2$ and $f_2(t) = 1 - t$. Define an inner product in terms of an integral as described above over the interval $[0, 1]$.
Suppose we want to find an orthogonal basis for Span $\{f_1, f_2\}$.
This is basically the Gram-Schmidt procedure. We want two new vectors (or functions) g_1 and g_2 that are orthogonal and span the same space as f_1 and f_2.
We begin by letting $g_1 = f_1$ and then define

$$\begin{aligned}
g_2 &= f_2 - \frac{\langle f_2, g_1\rangle}{\langle g_1, g_1\rangle} g_1 \\
&= 1 - t - \frac{\int_0^1 (t^2 - t^3)\, dt}{\int_0^1 t^4 \, dt} t^2 \\
&= 1 - t - \frac{1/12}{1/5} t^2 \\
&= 1 - t - \frac{5}{12} t^2
\end{aligned}$$

We can confirm that g_1 and g_2 are orthogonal

$$\begin{aligned}
\langle g_1, g_2 \rangle &= \int_0^1 t^2(1 - t - \frac{5}{12}t^2)\, dt \\
&= \int_0^1 (t^2 - t^3 - \frac{5}{12}t^4)\, dt \\
&= \frac{1}{3}t^3 - \frac{1}{4}t^4 - \frac{1}{12}t^5 \Big]_0^1 \\
&= \frac{1}{3} - \frac{1}{4} - \frac{1}{12} \\
&= 0
\end{aligned}$$

We will look a bit further at the definition of an inner product in terms of an integral. If you take the interval $[a, b]$ and divide it into an evenly spaced set of subintervals each having width δt then you should recall from calculus that

$$\int_a^b f(t)g(t)\, dt \approx \delta t \sum f(t_i)g(t_i)$$

where the sum is taken over all the right or left hand endpoints of the subintervals. But the expression on the right is just the inner product of the two vectors resulting from sampling the functions f and g at the (righthand or lefthand) endpoints of the subintervals with a scaling factor of δt (the width of the subintervals). Equivalently you can look at the terms on the right hand side as samples of the function $f(t)g(t)$. So you can look at the inner product defined in terms of the integral as a limiting case of the inner product defined in terms of sampling as the space between the samples approaches 0.

Example 5.1.8

In $C\,[0, 1]$ let $f(t) = t$ and $g(t) = t^2$.
Using an inner product defined in terms of the integral we would get

$$\langle f, g \rangle = \int_0^1 t^3\, dt = .25$$

Sampling the functions at 1/3, 2/3, and 1 and taking the dot product we would get

$$(1/3)(1/9) + (2/3)(4/9) + (1)(1) = 4/3$$

Scaling this by the interval width we get $(1/3)(4/3) = 4/9 \approx .4444$. This value would be the area of the rectangles in **Figure 5.1**.
This type of picture should be familiar to you from calculus. The integral evaluated above gives the area under the curve t^3 from $t = 0$ to $t = 1$. The discrete inner product gives an approximation to this area by a set of rectangles.

If we sample the functions at $0.1, 0.2, \ldots, 1.0$ and take the dot product we get

$$\sum_{i=1}^{10} f(i/10)g(i/10) = \sum_{i=1}^{10} i^3/1000 = 3.025$$

Scaling this by the interval width would give .3025, a result that is closer to the integral.
Figure 5.2 illustrates this approximation to the integral.

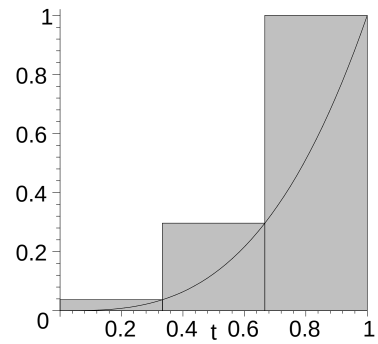


Figure 5.1: Sampling t^3 at 3 points.

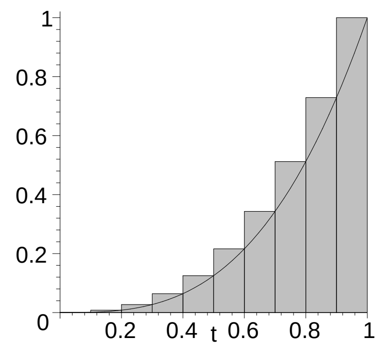


Figure 5.2: Sampling t^3 at 10 points.

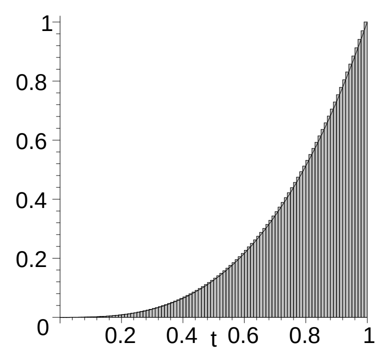


Figure 5.3: Sampling t^3 at 100 points.

If we sampled using an interval width of .01 the corresponding result would be .25050025. This result that is very close to the integral inner product.

Exercises

1. In $\mathbb{R}^2$ define the weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix} \mathbf{v}$.

   (a) Describe all vectors orthogonal to $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ in this inner product space.

   (b) Show that $\mathbf{u} = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ and $\mathbf{v} = \begin{bmatrix} -1 \\ 2 \end{bmatrix}$ are orthogonal in this inner product space and verify that $\|\mathbf{u}\|^2 + \|\mathbf{v}\|^2 = \|\mathbf{u} + \mathbf{v}\|^2$.

2. In $\mathbb{R}^2$ define the weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mathbf{v}$ where $a > 0$ and $b > 0$. Find the angle between $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ -1 \end{bmatrix}$ in terms of a and b relative to this inner product.

3. In $\mathbb{R}^2$ define the weighted inner product $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^T \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix} \mathbf{v}$ where $a > 0$ and $b > 0$.

   (a) Let
   $$\mathbf{u}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ \mathbf{u}_2 = \begin{bmatrix} 2 \\ -3 \end{bmatrix}$$
   Try to find specific weights a and b such that $\mathbf{u}_1$ and $\mathbf{u}_2$ will be orthogonal relative to the weighted inner product.

   (b) Let
   $$\mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \ \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$
   Try to find specific weights a and b such that $\mathbf{v}_1$ and $\mathbf{v}_2$ will be orthogonal relative to the weighted inner product.

4. In P_2 with $\langle p, q \rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$ let
   $$f(t) = 1 + t, \quad g(t) = t - t^2, \quad h(t) = t^2 + 2t - 2$$

   (a) Find $\langle f, g \rangle$

   (b) Find $\langle 2f - g, f + g \rangle$

   (c) Find $\|f\|$

   (d) Find the projection of g onto h.

   (e) Verify the Cauchy-Schwarz inequality for f and g. That is, verify that $|\langle f, g \rangle| \leq \|f\| \, \|g\|$.

   (f) Verify that f and h are orthogonal in this inner product space and that the Pythagorean Theorem, $\|f + h\|^2 = \|f\|^2 + \|h\|^2$, is satisfied by these vectors.

5. In P_2 define $\langle p, q \rangle = p(0)q(0) + p(1)q(1) + p(2)q(2)$. Let
   $$p(t) = t^2 + 2t + 1, \quad q(t) = t^2 + t$$

   (a) Find $\|p\|$.

   (b) Find the projection of p onto q and the orthogonal complement of this projection.

   (c) Find an orthogonal basis for the subspace of P_2 spanned by p and q.

(d) For what value(s) of α is $t+\alpha$ orthogonal to p.

6. In P_2 define $\langle p,q\rangle = p(-1)q(-1)+p(0)q(0)+p(1)q(1)$. Let W be the subspace of P_2 spanned by $f(t)=t$. Find a basis for $W^{\perp}$ in this inner product space.

7. In P_3 define $\langle p,q\rangle = p(-1)q(-1)+p(0)q(0)+p(1)q(1)+p(2)q(2)$. Let

$$p_0(t)=1, \quad p_1(t)=t, \quad p_2(t)=t^2, \quad p_3(t)=t^3$$

Let $W = \mathrm{Span}\,\{p_0,p_1,p_2\}$.

(a) Find an orthogonal basis for W.

(b) Find the best approximation to p_3 in W. Call this best approximation $\hat{p_3}$.

(c) Find $\|p_3-\hat{p_3}\|$.

(d) Verify that $p_3-\hat{p_3}$ is orthogonal to p_3.

8. In $C[0,1]$ with $\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt$ let $f(t)=\dfrac{1}{1+t^2}$ and $g(t)=2t$. Find

(a) $\langle f,g\rangle$

(b) $\|f\|$

(c) the projection of g onto f.

9. In $C[0,1]$ define $\langle f,g\rangle = \int_0^1 f(t)g(t)\,dt$. Convert the set of vectors $f_0(t)=1$, $f_1(t)=t$, $f_2(t)=t^2$, $f_3(t)=t^3$, into an orthogonal set by the Gram-Schmidt procedure.

10. In $C[-1,1]$ define $\langle f,g\rangle = \int_{-1}^1 f(t)g(t)\,dt$. Let

$$f(t)=t\,, \qquad g(t)=e^t+e^{-t}\,, \qquad h(t)=e^t-e^{-t}$$

(a) Find the projection of g onto f.

(b) Verify that g and h are orthogonal.

(c) Find the projection of f onto Span $\{g,h\}$.

11. In $C[0,\pi]$ with $\langle f,g\rangle = \int_0^\pi f(t)g(t)\,dt$ find an orthonormal basis for the

$$\text{Span } \left\{1,\sin t,\sin^2 t\right\}$$

12. In $M_{2\times 2}$ (the vector space of 2×2 matrices) an inner product can be defined as

$$\langle A,B\rangle = \mathrm{trace}(A^TB)$$

Let

$$A=\begin{bmatrix}1&1\\1&1\end{bmatrix},\ B=\begin{bmatrix}2&0\\1&1\end{bmatrix}$$

(a) Compute $\langle A,B\rangle$ and $\langle B,A\rangle$. Verify that these are the same.

(b) Find $\|A-B\|$.

(c) Find the projection of B onto A.

(d) Find a vector (i.e., matrix) orthogonal to A.

(e) Let W be the subspace of symmetric matrices. Show that $W^\perp$ is the set of skew-symmetric matrices.

13. In $C[-a,a]$ define $\langle f, g\rangle = \int_{-a}^{a} f(t)g(t)\,dt$.

    (a) Recall that a function is said to be **even** if $f(-t) = f(t)$ and a function is said to be **odd** if $f(-t) = -f(t)$. Show that if $f(t)$ is *any* function defined on $[-a,a]$ then $f(t) + f(-t)$ is even, and that $f(t) - f(-t)$ is odd.

    (b) Show that any function, $f(t)$, can be written as $f(t) = f_e(t) + f_o(t)$ where $f_e(t)$ is even and $f_o(t)$ is odd.

    (c) Show that if $f(t)$ is an even function in $C[-a,a]$ and $g(t)$ is an odd function in $C[-a,a]$ then $\langle f, g\rangle = 0$.

    (d) Write $t^4 + 3t^3 - 5t^2 + t + 2$ as the sum of an even function and an odd function.

    (e) Write e^t as the sum of an even and an odd function.

14. In $\mathbb{R}^n$ with the standard inner product show that $\langle \mathbf{u}, A\mathbf{v}\rangle = \langle A^T\mathbf{u}, \mathbf{v}\rangle$.

15. Let Q be an orthogonal $n \times n$ matrix and let $\mathbf{x}$ be any vector in $\mathbb{R}^n$.

    (a) Show that the vectors $Q\mathbf{x} + \mathbf{x}$ and $Q\mathbf{x} - \mathbf{x}$ are orthogonal relative to the standard inner product.

    (b) Show that the matrics $Q + I$ and $Q - I$ are orthogonal relative to the matrix inner product $\langle A, B\rangle = \mathrm{trace}(A^TB)$.

16. Let A be an $m \times n$ matrix with linearly independent columns. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Show that $\langle \mathbf{u}, \mathbf{v}\rangle = (A\mathbf{u})^TA\mathbf{v}$ defines an inner product.

17. Let A be a symmetric, positive definite, $n \times n$ matrix. Show that $\langle \mathbf{u}, \mathbf{v}\rangle = \mathbf{u}^TA\mathbf{v}$ defines an inner product for any two vectors $\mathbf{u}$ and $\mathbf{v}$ in $\mathbb{R}^n$.

18. Prove that the Pythagorean Theorem is true in any inner product space. That is, show that $\|\mathbf{u} + \mathbf{v}\|^2 = \|\mathbf{u}\|^2 + \|\mathbf{v}\|^2$ if and only if $\langle \mathbf{u}, \mathbf{v}\rangle = 0$.

19. Explain why $\|\mathbf{u} - \mathbf{v}\| = \|\mathbf{v} - \mathbf{u}\|$ in any inner product space. That is, explain why this equation is a consequence of the four conditions that define an inner product space.

20. Let $\mathbf{u}$ and $\mathbf{v}$ be vectors in $\mathbb{R}^n$. Let B be an $n \times n$ matrix whose columns form a basis $\mathcal{B}$ of $\mathbb{R}^n$. How could you define an inner product on $\mathbb{R}^n$ such that the dot product of $[\mathbf{u}]_\mathcal{B}$ and $[\mathbf{v}]_\mathcal{B}$ gives the same result as $\mathbf{u} \cdot \mathbf{v}$.

Using MAPLE

Example 1.

We will define an inner product in P_5 by sampling at $-3, -2, -1, 1, 2, 3$ and then taking the dot product of the resulting vectors. When we define the polynomials **Maple** gives us two options. They can be defined as expressions or functions. In this case it will be simpler to define them as functions.

We will define the polynomials

$$\begin{aligned} p_1(x) &= 1 + x + x^2 + x^3 + x^4 + x^5 \\ p_2(x) &= 2 - 2x + x^2 - x^3 + 2x^4 - 2x^5 \end{aligned}$$

and then illustrate various computations using **Maple** . Note that we use the symbol % at a couple points. This symbol refers to the output of the immediately previously executed **Maple** command.

```
>p1:=x->1+x+x^2+x^3+x^4+x^5;
>p2:=x->2-2*x+x^2-x^3+2*x^4-2*x^5;
>xv:=[-3,-2,-1,1,2,3]:
>v1:=map(p1,xv):  ## sampling p1
>v2:=map(p2,xv):   ## sampling p2
>v1^%T.v2:   ## the inner product of p1 and p2
                -256676
>sqrt(DotProduct(v1,v1)): evalf(%);  ## the magnitude of p1
                 412.3905916
>DotProduct(v1,v2)/sqrt(DotProduct(v1,v1))/sqrt(DotProduct(v2,v2)):
>evalf(arccos(%));   ## the angle between p1 and p2
                              2.489612756
>p3:=DotProduct(v1,v2)/DotProduct(v2,v2)*p2: ## the projection of p1 onto p2
>p4:=p1-p3:  ## the orthogonal complement
>v4:=map(p4,xv):   ## sample p4
>DotProduct(v4,v2);    ## these should be orthogonal. Are they?
               0
```

Example 2.

In this example we will look at $C\,[-1,1]$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$. We will look at the two functions $f = \cos(x)$ and $g = \cos(x+k)$ and plot the angle between these functions for different values of k. In this case it will be simpler to define f and g as expressions. The third line defines a **Maple** procedure called ip which requires two inputs and will compute the inner product of those inputs.

```
>f:=cos(x);
>g:=cos(x+k):
>ip:= (u,v) -> int( u*v, x=-1..1):  ### we define our inner product
>ang:=arccos( ip(f,g)/sqrt(ip(f,f))/sqrt(ip(g,g)) ):
>plot([ang, Pi/2], k=-4..12);
>solve(ang=Pi/2, k);
```

In our plot we included the plot of a horizontal line (a constant function) at $\pi/2$. The points where the plots intersect correspond to values of k which make the functions orthogonal. The last line shows that the first two such points occur at $k = \pm\pi/2$.

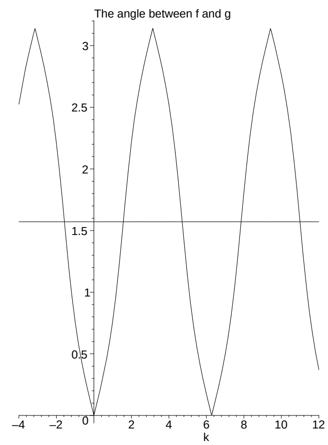


Figure 5.4:

The plot should make sense. When $k = 0$ we would have $f = g$ so the angle between them is 0. If $k = \pi$ it means the cosine funtion is shifted by π radians, and this results in $g = -\cos(x) = -f$ so the angle between f and g should be π radians. If $k = 2\pi$ the cosine will be shifted one complete cycle so we will have $f = g$ and so the angle between them will again be 0. This pattern will continue periodically.

Example 3.

Suppose we want to find the projection of $f = \sin(t)$ onto $g = t$ in the inner product space $C[-\pi/2, \pi/2]$ with

$$\langle f, g\rangle = \int_{-\pi/2}^{\pi/2} f(t)g(t)\,dt$$

In **Maple** we could enter the following commands:

```
>f:=sin(t):
>g:=t:
>ip:=(u,v)->int(u*v,t=-Pi/2..Pi/2): ## the inner product procedure
>proj:=ip(f,g)/ip(g,g)*g;
```

This gives us the projection $\frac{24}{\pi^3}t$.

Another way of looking at this is the following. The vector g spans a subspac e of $C[-\pi/2, \pi/2]$ consisting of all functions of the form kt (i.e., all straight lines through the origin). The projection of f onto g is the function of the form kt that is closest to g in this inner product space. The square of the distance from kt to g would be $\int_{-\pi/2}^{\pi/2} [kt - \sin(t)]^2\, dt$. We want to minimize this. In **Maple** :

```
>d:=int((k*t-sin(t))^2,t=-Pi/2..Pi/2); ## the distance squared
>d1:=diff(d,k);   ## take the derivative to locate the minimum
>solve(d1=0,k);   ## find the critical value

                k = 24/Pi^3

>d2:=diff(d1,k);  ## for the 2nd derivative test
```

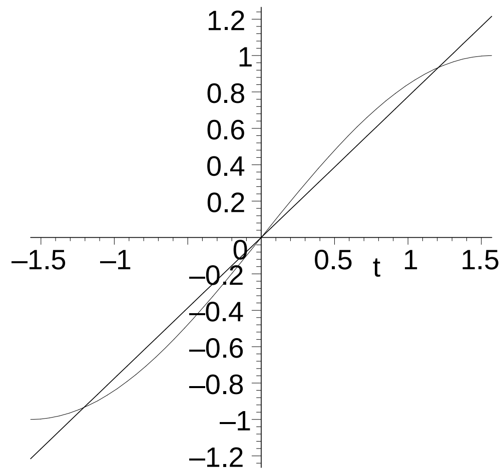


Figure 5.5: Plot of $\sin t$ and $(24/\pi^3)t$

Here we used the techniques of differential calculus to find out where the distance is a minimum (the second derivative test confirms that the critical value gives a minimum). We got the same result as before: $k = \dfrac{24}{\pi^3}$.

We will look at this same example a little further. We just found the projection of one function onto another in the vector space $C[0, 1]$, but this can be approximated by discrete vectors in $\mathbb{R}^n$ if we sample the functions. We will use **Maple** to sample the functions 41 times over the interval $[-\pi/2, \pi/2]$ and then find the projection in terms of these vectors in $\mathbb{R}^{41}$.

```
>f:=t->sin(t);  ## redefine f and g as functions, this makes things simpler
>g:=t->t;
>h:=Pi/40:   # the distance between samples
>xvals:=Vector(41, i->evalf(-Pi/2 + h*(i-1))):  # the sampling points
>u:=map(f,xvals):   ## the discrete versions of f and g
>v:=map(g,xvals):
>proju:=DotProduct(u,v)/DotProduct(v,v)*v:  # the projection of u onto v
```

We ended the last line above with a colon which means **Maple** doesn't print out the result. If you did see the result it would just be a long list of numbers and wouldn't mean much in that form. We will use graphics to show the similarity with out earlier result. We will plot the projection using the entries as y coordinates and the sampling points as the x coordinates.

```
>data:=[seq([xvals[i],proju[i]],i=1..41)]:
>p1:=plot(data,style=point):   # the discrete projection
>p2:=plot(24/Pi^3*t,t=-Pi/2..Pi/2):  # the continuous projection
>plots[display]([p1,p2]);
```

The plot clearly shows the similarity between the discrete and continuous projections. As a problem redo this example with only 10 sample points. Using fewer sample points should result in a greater disparity between the discrete and continuous cases.

Example 4.

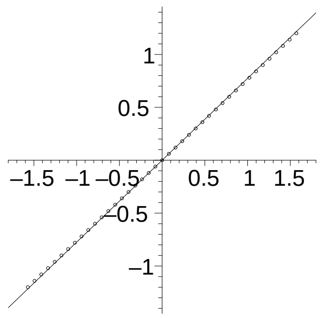


Figure 5.6: The discrete and continuous projections.

In $C[-1,1]$ with

$$\langle f,g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$$

the polynomials $\{1, t, t^2, t^3, \ldots t^8\}$ do not form an orthogonal set but we can apply the Gram-Schmidt procedure to convert them to an orthogonal basis. The integration required would be tedious by hand but **Maple** makes it easy:

```
>for i from 0 to 8 do f[i]:=t^i od;  # the original set
>ip:=(f,g)->int(f*g,t=-1..1):       # the inner product
>g[0]:=1:               # the first of the orthogonal set
>for i to 8 do
   g[i]:=f[i]-add(ip(f[i],g[j])/ip(g[j],g[j])*g[j],j=0..(i-1)) od:
```

The last line above might look complicated at first but it is just the way you would enter the basic Gram-Schmidt formula

$$g_i = f_i - \sum_{j=0}^{i-1} \frac{\langle f_i, g_j\rangle}{\langle g_j, g_j\rangle} g_j$$

into **Maple** .

This then gives us the orthogonal polynomials:

$$\begin{aligned}
g_0 &= 1\\
g_1 &= t\\
g_2 &= t^2 - \frac{1}{3}\\
g_3 &= t^3 - \frac{3}{5}t\\
g_4 &= t^4 - \frac{6}{7}t^2 + \frac{3}{35}\\
g_5 &= t^5 - \frac{10}{9}t^3 + \frac{5}{21}t\\
g_6 &= t^6 - \frac{15}{11}t^4 + \frac{5}{11}t^2 - \frac{5}{231}
\end{aligned}$$

These particular orthogonal polynomials are called the **Legendre polynomials**. They are useful precisely because they are orthogonal and we will be using them later in this chapter.

Here are the plots of the Legendre polynomials $g_2, \ldots, g_6$ and the standard basis functions $f_2, \ldots, f_6$.

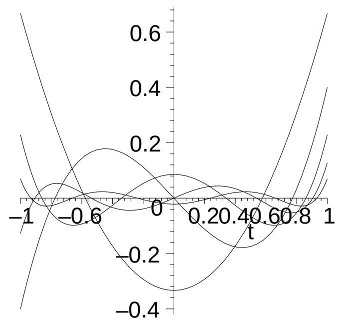


Figure 5.7: Legendre basis functions $g_2, \ldots, g_6$.

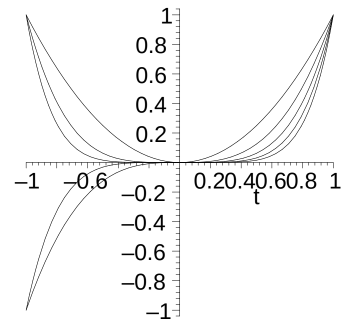


Figure 5.8: Standard basis functions $t^2, \ldots, t^6$.

These plots illustrate one drawback of the standard basis: for higher powers they become very similar. There is not much difference between the plot of t^4 and t^6 for example. This means that it if you want to represent an arbitrary function in terms of this basis it can become numerically difficult to separate the components of these higher powers.

5.2 Approximations in Inner Product Spaces

We saw in the previous chapter that the operation of orthogonally projecting a vector **v** onto a subspace W can be looked on as finding the best approximation to **v** by a vector in W. This same idea can be now extended to any inner product space.

Example 5.2.9

I n P_2 with the inner product

$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$

let $W = \text{Span}\,\{1, t\}$ and $f(t) = t^2$. Find the best approximation to f by a vector in W. Note that vectors in W are just linear functions, so this problem can be seen as trying to find the linear function that gives the best approximation to a quadratic relative to this particular inner product.

First we have $\langle 1, t\rangle = (1)(-1) + (1)(0) + (1)(1) = 0$ so in fact $\{1, t\}$ is an *orthogonal* basis for W. The best approximation to f is just the orthogonal projection of f into W and this is given by

$$\hat{f} = \frac{\langle 1, f\rangle}{\langle 1, 1\rangle}1 + \frac{\langle t, f\rangle}{\langle t, t\rangle}t = \frac{\langle 1, t^2\rangle}{\langle 1, 1\rangle}1 + \frac{\langle t, t^2\rangle}{\langle t, t\rangle}t = \frac{2}{3}1 + \frac{0}{2}t = \frac{2}{3}$$

So the best linear approximation to t^2 is in fact the horizontal line $\hat{f} = 2/3$. Here is a picture of the situation:

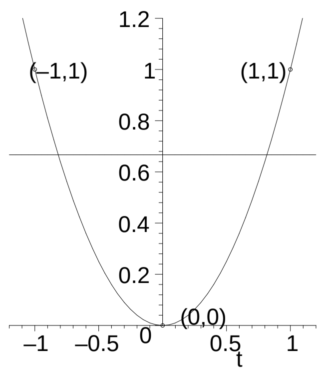


Figure 5.9: $f = t^2$ and $\hat{f} = 2/3$

In what sense is this line the best approximation to the quadratic t^2? One way of making sense of this is that the inner product used here only refers to the points at -1, 0, and 1. You can then look at this problem as trying to find the straight line that comes closest to the points (-1,1), (0,0), and (1,1). This is just the least-squares line and would be the line $y = 2/3$.

Now try redoing this example using the weighted inner product

$$\langle p, q\rangle = p(-1)q(-1) + \alpha p(0)q(0) + p(1)q(1)$$

where $\alpha > 0$. What happens as $\alpha \to 0$? What happens as $\alpha \to \infty$?

Example 5.2.10

Let W and f be the same as in the last example but now suppose that

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$$

With this inner product 1 and t are still orthogonal so the problem can be worked out the same as before. The only difference is

$$\langle 1, t^2\rangle = \int_{-1}^{1} t^2\,dt = \frac{2}{3} \qquad \langle 1, 1\rangle = \int_{-1}^{1} 1\,dt = 2 \qquad \langle t, t^2\rangle = \int_{-1}^{1} t^3\,dt = 0$$

so we get

$$\hat{f} = \frac{2/3}{2}1 = \frac{1}{3}$$

So in this inner product space the best linear approximation to t^2 would be the horizontal line $\hat{f} = 1/3$.

Example 5.2.11

Consider the inner product space $C[-1, 1]$ with inner product

$$\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$$

Let $W = \text{Span}\left\{1, t, t^2, t^3\right\}$ and $f(t) = e^t$. The problem is to find the best approximation to f in W relative to the given inner product. In this case the functions 1,t, t^2, and t^3 are not orthogonal but they can be converted to an orthogonal basis by the Gram-Schmidt procedure giving 1, t, $t^2 - 1/3$, $t^3 - 3/5t$ (as we saw before, in one of the **Maple** examples from the last section, these are called *Legendre polynomials*).
The projection of f onto W will then be given by

$$\hat{f} = \frac{\int_{-1}^{1} e^t\,dt}{\int_{-1}^{1} 1\,dt}1 + \frac{\int_{-1}^{1} te^t\,dt}{\int_{-1}^{1} t^2\,dt}t$$

$$+\frac{\int_{-1}^{1}(t^2 - 1/3)e^t\,dt}{\int_{-1}^{1}(t^2 - 1/3)^2\,dt}(t^2 - 1/3) + \frac{\int_{-1}^{1}(t^3 - 3/5t)e^t\,dt}{\int_{-1}^{1}(t^3 - 3/5t)^2\,dt}(t^3 - 3/5t)$$

It's not recommended that you do the above computations by hand. Using math software to evaluate the integrals and simplifying we get

$$\hat{f} = 0.9962940209 + 0.9979548527\,t + 0.5367215193\,t^2 + 0.1761391188\,t^3$$

Figure 5.10 is a plot of f and $\hat{f}$;
Notice the graphs seem to be almost indistinguishable on the interval $[-1, 1]$. How close is $\hat{f}$ to f? In an inner product space this means: what is $\|f - \hat{f}\|$? In this case we get

$$\|f - \hat{f}\|^2 = \int_{-1}^{1} (f - \hat{f})^2\,dt = .00002228925$$

So taking the square root we get $\|f - \hat{f}\| = .004721149225$
This example can be seen as trying to find a good polynomial approximation to the exponential function e^t. From calculus you should recall that another way of finding a

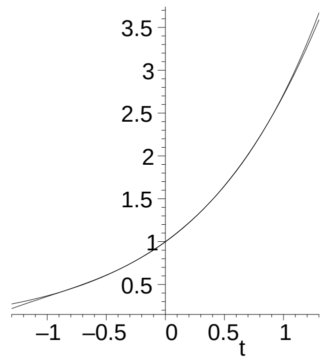


Figure 5.10: Plots of f and $\hat{f}$.

polynomial approximation is the Maclaurin series and that the first four terms of the Maclaurin series for e^t would be

$$g = 1 + t + \frac{1}{2}t^2 + \frac{1}{6}t^3 = 1 + t + .5t^2 + .1666666667t^3$$

How far is this expression from f in our inner product space? Again we compute $\|f - g\|^2 = \int_{-1}^{1}(f-g)^2\,dt$ and take the root which gives $\|f - g\| = .02050903825$ which is not as good as the projection computed first. That should not be surprising because in this inner product space *nothing* could be better than the orthogonal projection.
Figure 5.11 shows plots of $f - \hat{f}$ and $f - g$. These plots allow you to see the difference between f and the two polynomial approximations.

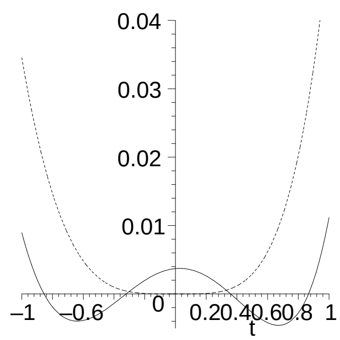


Figure 5.11: Plots of $f - \hat{f}$ (solid line) and $f - g$ (dotted line).

Notice that the Maclaurin series gives a better approximation near the origin (where this power series is centered) but at the expense of being worse near the endpoints. This is typical of approximations by Taylor or Maclaurin series. By contrast, the least-squares approximation tries to spread the error out evenly over the entire interval.

Exercises

1. In P_2 with the inner product
$$\langle p, q\rangle = p(-1)q(-1) + p(0)q(0) + p(1)q(1)$$
let $p(t) = t^2 - 2t$.
   (a) Find the linear function that gives the best approximation to p in this inner product space. Plot p and this linear approximation on the same set of axes.
   (b) Find the line through the origin that gives the best approximation to p. Plot p and this line on the same set of axes.
   (c) Find the horizontal line that gives the best approximation to p. Plot p and this line on the same set of axes.
2. In $C[-1, 1]$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let $f(t) = t^3$.
   (a) Suppose you approximate $f(t)$ by the straight line $g(t) = t$. Plot f and g on the same set of axes. What is the error of this approximation?
   (b) Find the straight line through the origin that gives the best approximation to f. Plot f and this best approximation on the same set of axes. What is the error of this best approximation?
3. In $C[-1, 1]$ with the inner product $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Let $f(t) = |t|$. Use Legendre polynomials to solve the following problems.
   (a) Find the best quadratic approximation to f. Plot f and this approximation on the same set of axes. What is the error of this approximation.
   (b) Find the best quartic (fourth degree polynomial) approximation to f. Plot f and this approximation on the same set of axes. What is the error of this approximation.
4. In $C[0, 1]$ with the inner product $\langle f, g\rangle = \int_{0}^{1} f(t)g(t)\,dt$. Let $f(t) = t^3$ and let $g(t) = t$.
   (a) Plot $f(t)$ and $g(t)$ on the same set of axes.
   (b) Find $\|f - g\|$.
   (c) Find the best approximation to $f(t)$ by a straight line through the origin.
   (d) Find the best approximation to $f(t)$ by a line of slope 1. This can be interpreted as asking you to find out by how much should $g(t)$ be shifted vertically in order to minimize the distance from $f(t)$.
5. In $C[0, 1]$ with the inner product $\langle f, g\rangle = \int_{0}^{1} f(t)g(t)\,dt$. Let $f(t) = t^n$ where n is a positive integer.
   (a) Find the line through the origin that gives the best approximation to f in this inner product space.
   (b) Find the line of slope 1 that gives the best approximation to f in this inner product space.
6. In $C[a, b]$ with the inner product $\langle f, g\rangle = \int_{a}^{b} f(t)g(t)\,dt$:
   (a) Find an orthogonal basis for Span $\{1, t\}$.
   (b) Use the orthogonal basis just found to find the best linear approximation to $f(t) = t^2$ in this inner product space. Call this best approximation $\hat{f}$.
   (c) Evaluate $\int_{a}^{b} f(t)\,dt$ and $\int_{a}^{b} \hat{f}(t)\,dt$.
7. Let $\hat{f}$ be the best approximation to f in some inner product space and let $\hat{g}$ be the best approximation to g in the same space. Is $\hat{f} + \hat{g}$ the best approximation to $f + g$?

Using MAPLE

Example 1.

In the inner product space $C[-1,1]$ with $\langle f,g\rangle = \int_{-1}^{1} f(x)g(x)\,dx$ suppose you want to find the best approximation to $f(x)=e^x$ by a polynomial of degree 4. The set of functions $\{1, x, x^2, x^3, x^4\}$ is a basis for the subspace of polynomials of degree 4 or less, but this basis is not orthogonal in this inner product space. Our first step will be to convert them into an orthogonal basis. But this is just what we did in the last section when we found the **Legendre polynomials**. If you look back at that computation you will see that the orthogonal polynomials we want are

$$1, \quad x, \quad x^2 - \frac{1}{3}, \quad x^3 - \frac{3}{5}x, \quad x^4 - \frac{6}{7}x^2 + \frac{3}{35}$$

We will define these in **Maple**

```
>g[1]:=1:
>g[2]:=x:
>g[3]:=x^2-1/3:
>g[4]:=x^3-3/5*x:
>g[5]:=x^4-6/7*x^2+3/35:
```

Now we find the best approximation by projecting into the subspace

```
>f:=exp(x);
>ip:=(f,g)->int(f*g,x=-1..1):
>fa:=add(ip(f,g[i])/ip(g[i],g[i])*g[i],i=1..5);
>fa:=evalf(fa);
```

$$1.000031 + 0.9979549\,x + 0.4993519\,x^2 + 0.1761391\,x^3 + 0.04359785\,x^4$$

```
>plot([f,fa],x=-1..1);
>plot([f,fa],x=-1.1..-.9);
```

The first of these plots shows the polynomial approximation and e^x. They are almost indistinguishable over the given interval. The second plot, shown in **Figure 5.12**, shows some divergence between the the two functions near the left-hand endpoint.

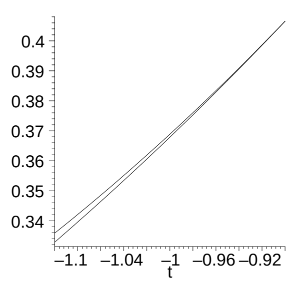


Figure 5.12: e^x and the polynomial approximation near $x=-1$.

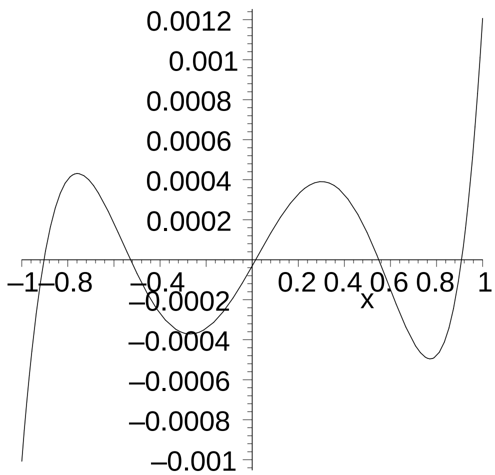


Figure 5.13: Plot of $f - \hat{f}$

Figure 5.13 shows the plot of $f - \hat{f}$. It shows clearly the difference between the best approximation we computed and f, and that the error is greatest at the endpoints.

It is also interesting to compare this approximation with the first 5 terms of the Taylor series for e^x. The Taylor polynomial would be

$$1.0000000 + 1.0000000x + .5x^2 + .1666667x^3 + .0416667x^4$$

These coefficients are very close to those of our least squares approximation but notice what happens when we compare distances. We will define `ft` to be the Taylor polynomial.

```
>ft:=convert(taylor(f,x=0,5),polynom):
>evalf(sqrt(ip(f-fa,f-fa)));
                              .0004711687596
>evalf(sqrt(ip(f-ft,f-ft)));
                              .003667974153
```

So in this inner product space `fa` is closer to `f` than `ft` by roughly a factor of 10.

Example 2.

In this example we will use **Maple** to find an polynomial approximation to $f(x) = x + |x|$ in the same inner product space as our last example. We will find the best approximation by a tenth degree polynomial. In this case we will need the Legendre polynomials up to the 10^{th} power. Fortunately **Maple** has a command which will generate the Legendre polynomials. The `orthopoly` package in **Maple** loads routines for various types of orthogonal polynomials (see the **Maple** help screen for more information). The command we are interested in is the single letter command `P` which produces Legendre polynomials.

```
>with(orthopoly);
>for i from 0 to 10 do g[i]:=P(i,x) od;
```

If you look at the resulting polynomials you will see they aren't exactly the same as the ones we computed, but they differ only by a scalar factor so their orthogonality is not affected.

Next we enter the function we are approximating. (You should plot this function if you don't know what it looks like.)

```
>f:=x+abs(x):
```

Now we apply the projection formula and plot the resulting approximation and plot the error of the approximation. In the projection formula the weight given to each basis function is defined as a ratio of two inner products. The first line below defines a procedure called `wt` which computes this ratio.

```
>wt:=(f,g)->int(f*g,x=-1..1)/int(g*g,x=-1..1):   ## define the weight
>fappr:=add(wt(f,g[i])*g[i],i=0..10):
>plot([f,fappr],x=-1.2..1.2,0..2.5);
>plot(f-fappr,x=-1..1);
```

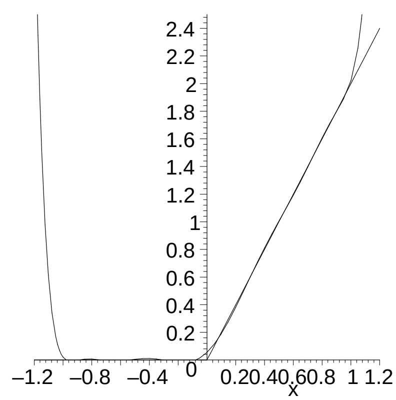


Figure 5.14: The plot of f and $\hat{f}$.

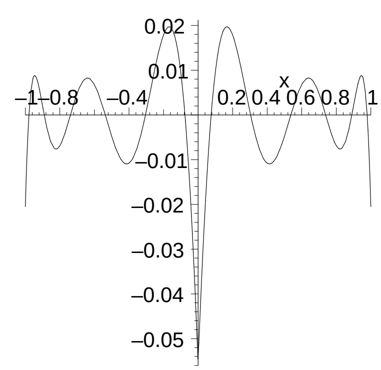


Figure 5.15: The plot of $f - \hat{f}$.

Notice that the approximation is worst near the origin where the original function is not differentiable. In this example it would be impossible to find a Taylor polynomial approximation to f on the given interval since f is not differentiable on the interval.

Example 3.

In our next example we will find a approximation to $f(x) = \frac{1}{1+x^2}$ in $C[0,1]$ with the inner product defined in terms of the integral. We will find the best approximation to f by a linear combination of the functions e^x and e^{-x}.

The first step is to get an orthogonal basis.

```
>f:=1/(1+x^2):
>f1:=exp(x):
>f2:=exp(-x):
>ip:=(f,g)->evalf(Int(f*g,x=0..1)): ## the evalf makes the results less messy
>wt:=(f,g)->ip(f,g)/ip(g,g):
>g1:=f1:   ## Gram-Schmidt step 1
>g2:=f2-wt(f2,g1)*g1;  ## Gram_Schmidt step 2
>fappr:=wt(f,g1)*g1 + wt(f,g2)*g2;
>plot([f,fappr],x=0..1);
```

We then get our best approximation as a linear combination of e^x and e^{-x}:

$$\hat{f} = 0.0644926796\, e^x + 1.064700466\, e^{-x}$$

But as the plot shows the approximation is not too good in this case.

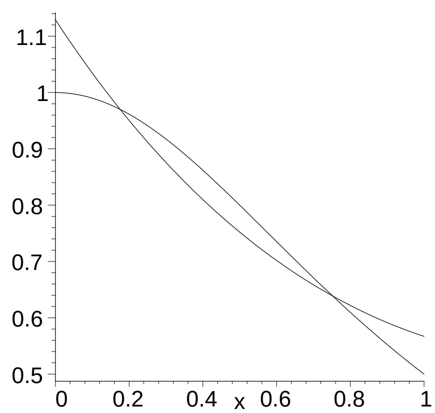


Figure 5.16: The plot of f and $\hat{f}$.

Example 4.

In this example we will illustrate a fairly complicated problem using **Maple** . We start with the function $f(t) = \frac{8t}{1+4t^2}$. We will sample this function at 21 evenly space points over the interval $[-1,1]$. Next we will find the polynomials of degree 1 to 20 that give the best least-squares fit to the sampled points. We will then find the distance from $f(t)$ to each of these polynomials relative to the inner product $\langle f, g\rangle = \int_{-1}^{1} f(t)g(t)\,dt$. Finally we will plot these distances versus the degree of the polynomials.

We start with some basic definitions we will need

```
>f:=t-> 8*t/(1+4*t^2):  ## define the function f
>xv:=convert( [seq(.1*i,i=-10..10)], Vector):  ## the x values
>yv:=map(f, xv):  ## the sampled values
>V:=VandermondeMatrix(xv):
```

Next we will define the best fitting polynomials which we will call p[1] to p[20]. We will use a loop to define these. Remember that the coefficients in each case are computed by finding the least-squares solution to a system where the coefficient matrix consists of columns of the Vandermonde matrix. We will use the QR decomposition to find the least-squares solution.

```
>for i from 2 to 21 do
   A:=V[1..21,1..i]:
   Q,R:=QRdecomposition(A, output=['Q','R']):
   sol:=R^(-1). Q^%T .yv):
   p[i-1]:=add(sol[j]*t^(j-1),j=1..i):
   od:
```

We now have our polynomials p[1], p[2],..., p[20]. We now want to compute the distance from $f(t)$ to each of these polynomials. We will call these values err[1] to err[20] and use a loop to compute them. First we will define the inner product and a command to find the length of a vector. Notice that f has been defined as a function but the polynomials have been defined as expressions.

```
>ip:=(f,g)->int(f*g,t=-1..1):
>lngth:=f->sqrt(ip(f,f)):
>for i from 1 to 20 do
 err[i]:=lngth( f(t)-p[i] );
 od:
>plot([seq( [i,err[i]],i=1..20)]);
```

The resulting plot is shown in **Figure 5.17**.

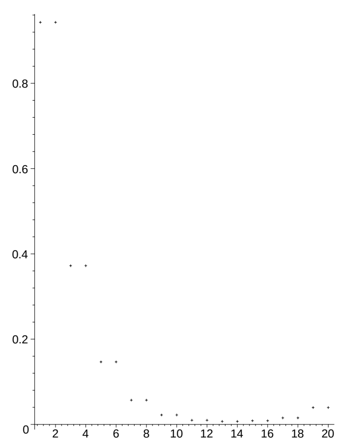


Figure 5.17: The distance from p[i] to f.

There are a couple of interesting aspects of this plot. First, notice that the points come in pairs. If you have a polynomial approximation of an odd degree then adding an additional even power term doesn't give you a better approximation. This is a consequence of the fact that $f(t)$ is an odd function and in this inner product space even and odd functions are orthogonal.

5.3 Fourier Series

We have seen that the Legendre polynomials form an orthogonal set of polynomials (relative to a certain inner product) that are useful in approximating functions by polynomials. There are many other sets of orthogonal functions but the most important basis consisting of orthogonal functions is the one we will look at in this section.

Consider the inner product space $C[-\pi,\pi]$ with $\langle f,g\rangle = \int_{-\pi}^{\pi} f(t)g(t)\,dt$. Notice that if m and n are integers then $\cos(mt)$ is an even function and $\sin(nt)$ is an odd function. By **Exercise 13** from section 5.1 this means that they are orthogonal in this inner product space. Moreover, if m and n are integers and $m \neq n$ then

$$\begin{aligned}
\langle \cos(mt), \cos(nt)\rangle &= \int_{-\pi}^{\pi} \cos(mt)\cos(nt)dt \\
&= \frac{1}{2}\int_{-\pi}^{\pi} \left[\cos(m+n)t + \cos(m-n)t\right]\,dt \\
&= \frac{1}{2}\left[\frac{\sin(m+n)t}{m+n} + \frac{\sin(m-n)t}{m-n}\right]_{-\pi}^{\pi} \\
&= 0
\end{aligned}$$

It follows that $\cos(mt)$ and $\cos(nt)$ are orthogonal for distinct integers m and n. A similar argument can be used to show that $\sin(mt)$ and $\sin(nt)$ are orthogonal for distinct integers m and n.

We can now conclude that $\mathcal{F} = \{1, \cos t, \sin t, \cos(2t), \sin(2t), \cos(3t), \sin(3t), \dots\}$ is an orthogonal set[2] of functions (vectors) in this inner product space. Do these vectors span $C[-\pi,\pi]$? That is, can any function in $C[-\pi,\pi]$ be written as a linear combination of these sines and cosines? The answer to this question leads to complications that we won't consider in this book, but it turns out that any "reasonable" function can be written as a combination of these basis functions. We will illustrate what this means with the next few examples.

Our goal is to find approximations to functions in terms of these orthogonal functions. This will be done, as before, by projecting the function to be approximated onto Span $\mathcal{F}$. The weight given to each basis function will be computed as a ratio of two inner products. Using calculus it is not difficult to show that for any positive integer m

$$\langle \sin(mt), \sin(mt)\rangle = \int_{-\pi}^{\pi} \sin^2(mt)\,dt = \pi$$

and

$$\langle \cos(mt), \cos(mt)\rangle = \int_{-\pi}^{\pi} \cos^2(mt)\,dt = \pi$$

Also we have

$$\int_{-\pi}^{\pi} 1\,dt = 2\pi$$

So if we define

$$\begin{aligned}
a_n &= \frac{\langle f, \cos(nt)\rangle}{\pi} \qquad \text{for } n = 0, 1, 2, \dots \\
b_n &= \frac{\langle f, \sin(nt)\rangle}{\pi} \qquad \text{for } n = 1, 2, 3, \dots
\end{aligned}$$

[2]Since this set is orthogonal, it must also be linearly independent. The span of these vectors must then be an infinite dimensional vector space.

we then have

$$f(t) \sim \frac{a_0}{2} + a_1 \cos t + b_1 \sin t + a_2 \cos 2t + b_2 \sin 2t + a_3 \cos 3t + b_3 \sin 3t + \cdots$$

The right hand side of the above expression is called the Fourier series of f, and the orthogonal set of sines and cosines is called the real Fourier basis. Here the $\sim$ only means that the right hand side is the Fourier series for $f(t)$.

You should see a parallel between Fourier series and Maclaurin (or Taylor) series that you've seen in calculus. With a Maclaurin series you can express a wide range of functions (say with variable t) as a linear combination of $1, t, t^2, t^3, \ldots$. With a Fourier series you can express a wide range of functions as a linear combination of sines and cosines. If you terminate the Fourier series after a certain number of terms the result is called a Fourier polynomial. More specifically, the n^{th} order Fourier polynomial is given by

$$\frac{a_0}{2} + a_1 \cos t + b_1 \sin t + \cdots + a_n \cos nt + b_n \sin nt$$

There is one other aspect of Fourier series that can be of help with some calculations. Functions of the form $\cos nt$ are even, and function of the form $\sin nt$ are odd. So the Fourier series of any even function will contain only cosine terms (including the constant term $a_0/2$), and the Fourier series of any odd function will contain only sine terms.

Example 5.3.12

Suppose we want to find the Fourier series for $f(t) = t$. Then

$$a_k = \int_{-\pi}^{\pi} t \cos(kt)\, dt = 0$$

since t is odd and $\cos(kt)$ is even. The series that we are looking for will contain only sine terms.

Next (using integration by parts) we get

$$\begin{aligned} b_k &= \frac{1}{\pi} \int_{-\pi}^{\pi} t \sin(kt)\, dt \\ &= \frac{1}{\pi} \left[\frac{-t\cos(kt)}{k} + \frac{\sin(kt)}{k^2} \right]_{-\pi}^{\pi} \\ &= \begin{cases} -2/k & \text{if } k \text{ is even} \\ 2/k & \text{if } k \text{ is odd} \end{cases} \end{aligned}$$

The Fourier series would then be

$$\begin{aligned} f(t) &\sim 2\sin(t) - \sin(2t) + \frac{2}{3}\sin(3t) - \frac{1}{2}\sin(4t) + \cdots \\ &= 2\sum_{k=1}^{\infty} \frac{(-1)^{k+1}}{k} \sin(kt) \end{aligned}$$

Notice that $f(t)$ and the Fourier series are *not* equal to each other on the interval $[-\pi, \pi]$. If we substitute $t = \pm\pi$ into the Fourier series we get 0, but $f(\pi) = \pi$ and $f(-\pi) = -\pi$. Surprisingly, however, they would be equal at every other point in the interval. This convergence is illustrated in **Figures 5.18-5.19**.

Finally, notice that if you take the Fourier series for $f(t)$ and substitute $t = \pi/2$ you can derive the formula

$$\frac{\pi}{4} = 1 - \frac{1}{3} + \frac{1}{5} - \frac{1}{7} + \cdots$$

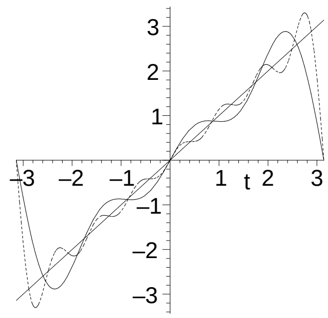


Figure 5.18: The plot of $f(t) = t$ and the Fourier approximations of order 3 and 7.

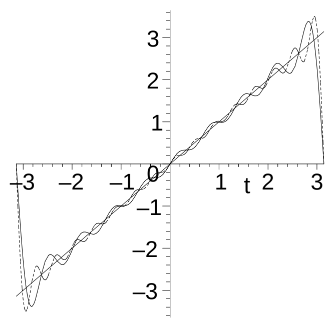


Figure 5.19: The plot of $f(t) = t$ and the Fourier approximations of order 9 and 15.

Example 5.3.13

N ext we will find the Fourier series for

$$f(t) = \begin{cases} 0 & \text{if } t \leq 0 \\ t & \text{if } t > 0 \end{cases}$$

This function is neither even nor odd and so the Fourier series will involve both sines and cosines. Computing the weights we get

$$\begin{aligned} a_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} f(t)\cos(kt)\,dt \\ &= \frac{1}{\pi}\int_{0}^{\pi} t\cos(kt)\,dt \\ &= \frac{1}{\pi}\left[\frac{t\sin(kt)}{k} + \frac{\cos(kt)}{k^2}\right]_0^{\pi} \\ &= \begin{cases} 0 & \text{if } k \text{ is even} \\ -2/(k^2\pi) & \text{if } k \text{ is odd} \end{cases} \end{aligned}$$

Similarly we get

$$b_k = \begin{cases} -1/k & \text{if } k \text{ is even} \\ 1/k & \text{if } k \text{ is odd} \end{cases}$$

A separate integration would give $a_0 = \pi/2$.
In this case the fifth order Fourier polynomial would be

$$\frac{\pi}{4} - \frac{2}{\pi}\cos(t) + \sin(t) - \frac{1}{2}\sin(2t) - \frac{2}{9\pi}\cos(3t)$$

$$+\frac{1}{3}\sin(3t)-\frac{1}{4}\sin(4t)-\frac{2}{25\pi}\cos(5t)+\frac{1}{5}\sin(5t)$$

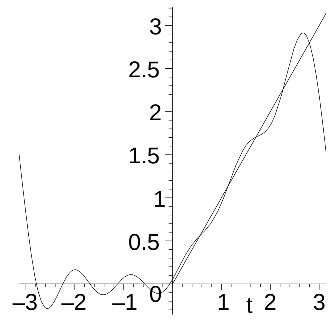


Figure 5.20: The plot of $f(t)$ and the Fourier approximation of order 5.

There is one more important aspect of a Fourier series. Suppose you have a sine and cosine wave at the same frequency but with different amplitudes. Then we can do the following:

$$\begin{aligned} A\cos(kt)+B\sin(kt) &= \sqrt{A^2+B^2}\left(\frac{A}{\sqrt{A^2+B^2}}\cos(kt)+\frac{B}{\sqrt{A^2+B^2}}\sin(kt)\right) \\ &= \sqrt{A^2+B^2}\,(\cos(\phi)\cos(kt)+\sin(\phi)\sin(kt)) \\ &= \sqrt{A^2+B^2}\cos(kt-\phi) \end{aligned}$$

This tells us that the combination of the sine and cosine waves is equivalent to one cosine wave at the same frequency with a phase shift and, more importantly, with an amplitude of $\sqrt{A^2+B^2}$. In other words any Fourier series involving sine and cosine terms at various frequencies can always be rewritten so that it involves just one of these trig functions at any frequency with an amplitude given by the above formula.

Exercises

1. In $C[-\pi,\pi]$ find the Fourier series for the (even) function
$$f(t) = |t|$$
Plot $f(t)$ and the 6^{th} order Fourier polynomial approximation to $f(t)$ on the same set of axes.

2. In $C[-\pi,\pi]$ find the Fourier series for the (even) function
$$f(t) = 1 - t^2$$
Plot $f(t)$ and the 6^{th} order Fourier polynomial approximation to $f(t)$ on the same set of axes.

3. In $C[-\pi,\pi]$ find the Fourier series for the (odd) function
$$f(t) = \begin{cases} -1 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$
Plot $f(t)$ and the 8^{th} order Fourier polynomial approximation to $f(t)$ on the same set of axes.

4. Find the Fourier series for the function
$$f(t) = \begin{cases} 0 & \text{if } t < 0 \\ 1 & \text{if } t \geq 0 \end{cases}$$
Plot $f(t)$ and the 8^{th} order Fourier polynomial approximation to $f(t)$ on the same set of axes.

5. Sample the functions 1, $\cos t$, $\sin t$, and $\cos 2t$ at the values $t = \pi/4, 3\pi/4, 5\pi/4, 7\pi/4$. Do the sampled vectors give an orthogonal basis for $\mathbb{R}^4$.

6. (a) Plot the weight of the constant term in the Fourier series of $\cos(kt)$ for $0 \leq k \leq 2$.
   (b) Plot the weight of the $\cos(t)$ term in the Fourier series of $\cos(kt)$ for $0 \leq k \leq 2$.

Using MAPLE

Example 1.

In this section we saw that the Fourier series for $f(t) = t$ is given by

$$2\sum_{k=1}^{\infty}\frac{(-1)^{k+1}}{k}\sin(kt)$$

and it was claimed that this series and $f(t)$ were equal over $[-\pi, \pi]$ except at the end points of this interval. We will illustrate this by animating the difference between $f(t)$ and its Fourier polynomial approximation of order N. The first line of the **Maple** code below defines the difference between t and the Fourier polynomial approximation of order N as a function of N. So, for example, if you enter `err(3)` then **Maple** would return

$$2\sin(t) - \sin(2t) + \frac{2}{3}\sin(3t)$$

```
>err:=N -> t - 2*add((-1)^(k+1)*sin(k*t)/k,k=1..N):
>for i to 40 do p[i]:=plot(err(i),t=-Pi..Pi,color=black) od:
>plots[display]( [seq( p[i],i=1..40)],insequence=true);
```

Now if you play the animation you have a visual representation of the difference between $f(t)$ and its Fourier approximations. **Figure 5.21** shows the plots of the error for the approximations of orders 5 and 25. Notice the behavior at the end points. Even though the magnitude of the error seems to get smaller in the interior of the interval, the error at the endpoints seems to remain constant.

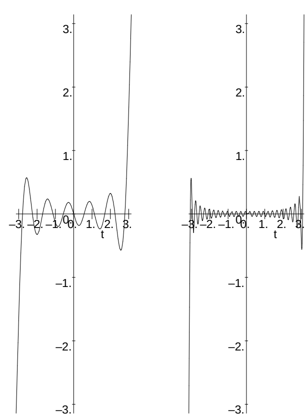


Figure 5.21: The error of Fourier approximations of order 5 and 25.

Example 2.

Let $f(x) = |x^2 - x|$. We will use **Maple** to compute Fourier approximations for this function. We begin by defining this function in **Maple** and plotting it.

```
>f:=abs(x^2-x);
>plot(f,x=-Pi..Pi);
```

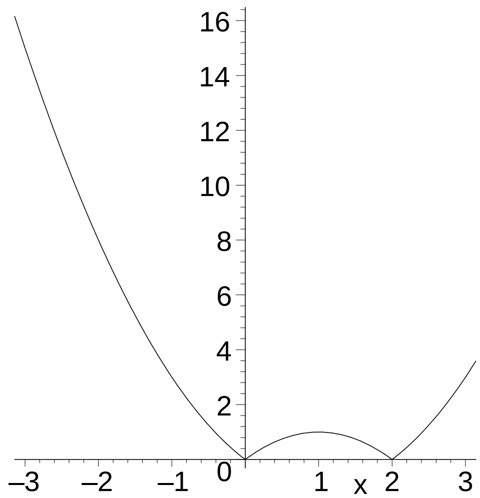


Figure 5.22: The plot of $f(x) = |x^2 - x|$.

This gives us the plot shown in **Figure 5.22**. Notice that this function is neither even nor odd and so the Fourier series should have both cosine and sine terms.

Next we have to compute the coefficients of the Fourier series. For the cosine terms we have $a_n = \dfrac{1}{\pi}\displaystyle\int_{-\pi}^{\pi} f(x)\cos(nx)\,dx$. In **Maple** we will compute the coefficients from a_0 to a_{20}:

```
>for i from 0 to 20 do a[i]:=evalf(int(f*cos(i*x),x=-Pi..Pi)/Pi) od;
```

Note that without the `evalf` in the above command **Maple** would have computed the exact values of the integrals which would result in very complicated expressions for the coefficients. Try it and see.

Next we find the coefficients of the sine functions in a similar way:

```
>for i from 1 to 20 do b[i]:=evalf(int(f*sin(i*x),x=-Pi..Pi)/Pi) od;
```

Now we can define the Fourier approximations of order 8 and of order 20 and plot them along with the original function:

```
>f8:=a[0]/2+add(a[i]*cos(i*x)+b[i]*sin(i*x),i=1..8);
>f20:=a[0]/2+add(a[i]*cos(i*x)+b[i]*sin(i*x),i=1..20);
>plot([f,f8],x=-Pi..Pi);
>plot([f,f20],x=-Pi..Pi);
```

This gives the plots shown in **Figures 5.23** and **5.24**.

To make things a bit more interesting try entering the following:

```
>for n to 20 do f[n]:=a[0]/2+add(a[i]*cos(i*x)+b[i]*sin(i*x),i=1..n) od;
>for n to 20 do p[n]:=plot([f,f[n]],x=-Pi..Pi) od:
>plots[display]([seq(p[n], n=1..20)],insequence=true);
```

The first command here defines all of the Fourier approximations from order 1 to order 20. The second command defines plots of these approximations (MAKE SURE YOU END THIS LINE WITH A COLON. Otherwise you'll get a lot of incomprehensible stuff printed out in you **Maple** worksheet.) The third command uses the `display` command from the `plots` package and creates an animation of the

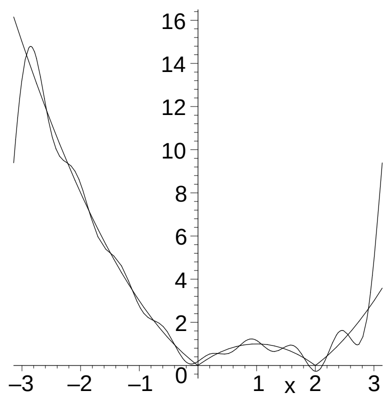


Figure 5.23: The order 8 Fourier approximation to f.

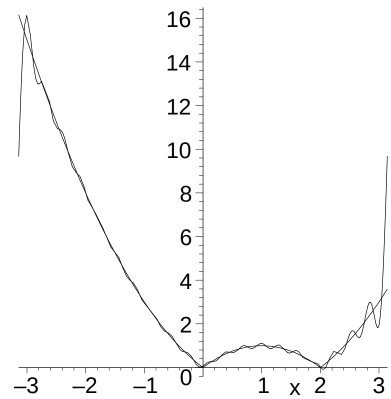


Figure 5.24: The order 20 Fourier approximation to f.

Fourier approximations converging to $f(x)$. Click on the plot that results, then click on the "play" button to view the animation.

Now go back and try the same steps several times using various other initial functions.

As a further exercise you could try animating the errors of the approximations.

Example 3.

We have seen that $\cos(t)$ and $\cos(k*t)$ are orthogonal for integer values of k $(k \neq 1)$ using the integral inner product over $[-\pi, \pi]$. What happens if k is not an integer? That is, what is the angle between $\cos(t)$ and $\cos(k*t)$ for any value of k? We also saw that the norm of $\cos(k*t)$ for non-zero integers k is $\sqrt{\pi}$. What is the norm for other values of k?

The following **Maple** commands will answer the questions posed above. The first line defines the relevant inner product. The fourth line is just the formula for finding the angle between vectors in an inner product space.

```
>ip:=(f,g)->int(f*g,t=-Pi..Pi):
>f1:=cos(t):
>f2:=cos(k*t):
>theta:=arccos( ip(f1,f2)/sqrt(ip(f1,f1)*ip(f2,f2)));
>plot( [theta,Pi/2],k=0..8);
>lgth:=sqrt(ip(f2,f2)):
>plot([lgth,sqrt(Pi)],k=0..8);
```

We get the plots shown in **Figures 5.25** and **5.26**.:

Figure 5.25 shows that the graph of the angle approaches $\pi/2$ as a horizontal asymptote, so $\cos(t)$ and $\cos(kt)$ are almost orthogonal for any large value of k. **Figure 5.26** shows that something similar happens with the norm. The graph of the norm approaches $\sqrt{\pi}$ as a horizontal asymptote, so for large values of k the function $\cos(kt)$ has a norm of approximately $\sqrt{\pi}$.

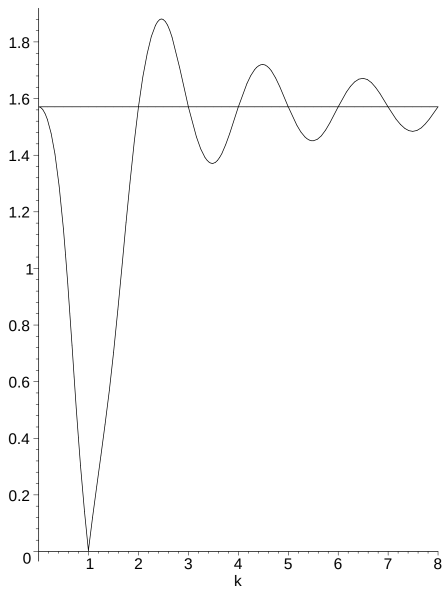


Figure 5.25: The angle between $\cos(t)$ and $\cos(kt)$.

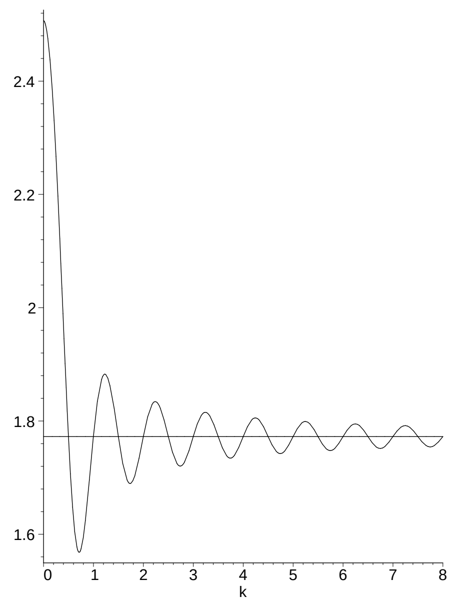


Figure 5.26: The norm of $\cos(kt)$.

5.4 Discrete Fourier Transform

When you find the Fourier series for some function $f(t)$ you are essentially approximating $f(t)$ by a linear combination of a set of orthogonal functions in an infinite dimensional vector space. But, as we have seen before, functions can be sampled and the sampled values constitute a finite dimensional approximation to the original function. Now if you sample the functions $\sin(mt)$ and $\cos(nt)$ (where m and n are integers) at equally spaced intervals from 0 to 2π the sampled vectors that result will be orthogonal (a proof of this will be given in the next chapter). A basis for $\mathbb{R}^n$ made from these sampled functions is called a Fourier basis[3].

For example, in $\mathbb{R}^4$, the Fourier basis would be

$$\mathbf{v}_1 = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}, \mathbf{v}_2 = \begin{bmatrix}1\\0\\-1\\0\end{bmatrix}, \mathbf{v}_3 = \begin{bmatrix}0\\1\\0\\-1\end{bmatrix}, \mathbf{v}_4 = \begin{bmatrix}1\\-1\\1\\-1\end{bmatrix}$$

What is the connection between these basis vectors and sines and cosines? Look at **Figures 5.27-5.30**.

Notice that this basis consists of cosine and sine functions sampled at 3 different frequencies. Notice also that sampling $\sin(0 \cdot t)$ and $\sin(2 \cdot t)$ gives the zero vector which would not be included in a basis.

[3]We will be sampling over the interval $[0, 2\pi]$ but we could also have used the interval $[-\pi, \pi]$. All we need is an interval of length 2π.

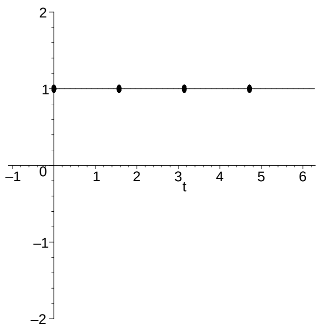


Figure 5.27: $\cos(0 \cdot t)$ with sampled points.

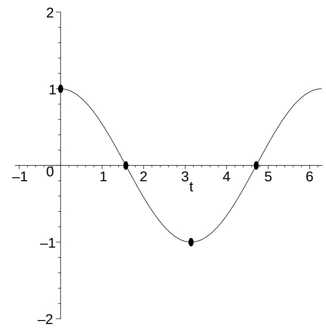


Figure 5.28: $\cos(1 \cdot t)$ with sampled points.

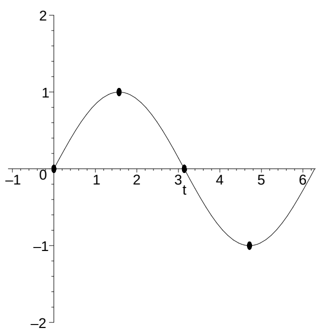


Figure 5.29: $\sin(1 \cdot t)$ with sampled points.

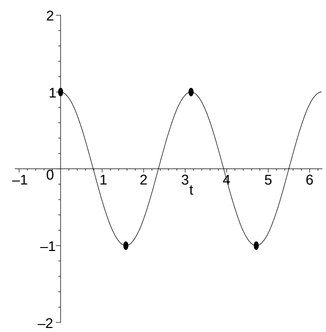


Figure 5.30: $\cos(2 \cdot t)$ with sampled points.

The Discrete Fourier basis for $\mathbb{R}^8$

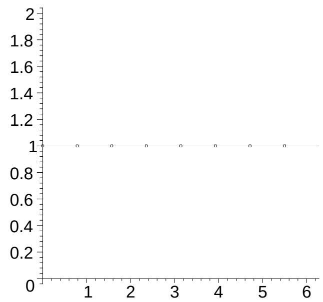


Figure 5.31: $\cos(0 \cdot t)$ sampled 8 times.The lowest frequency component: $[1, 1, 1, 1, 1, 1, 1, 1]$

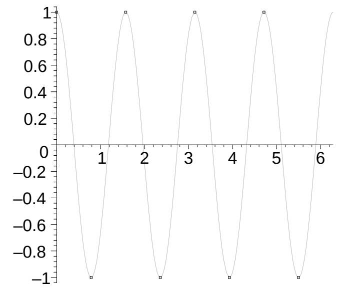


Figure 5.32: $\cos(4 \cdot t)$ sampled 8 times. The highest frequency component: $[1, -1, 1, -1, 1, -1, 1, -1]$

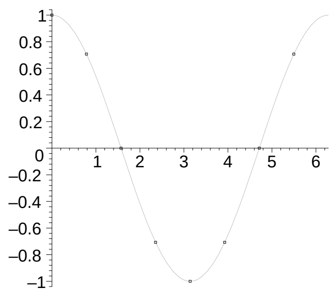


Figure 5.33: $\cos(1 \cdot t)$ with sampled points.

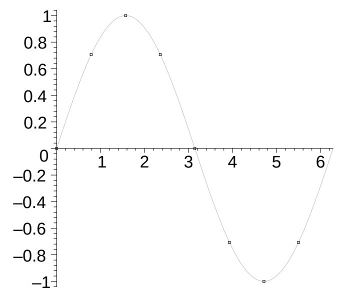


Figure 5.34: $\sin(t)$ with sampled points.

The Discrete Fourier basis for $\mathbb{R}^8$ (continued)

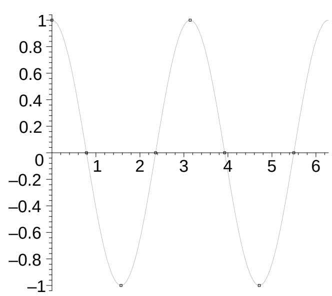


Figure 5.35: $\cos(2t)$ with sampled points.

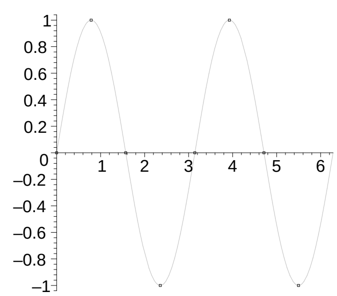


Figure 5.36: $\sin(2t)$ with sampled points.

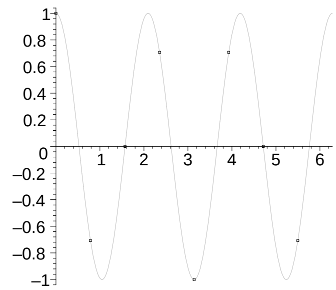


Figure 5.37: $\cos(3t)$ with sampled points.

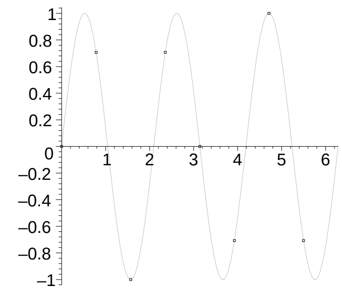


Figure 5.38: $\sin(3t)$ with sampled points.

Example 5.4.14

T he columns of the following matrix are the Fourier basis for $\mathbb{R}^8$.

$$F = \begin{bmatrix} 1 & 1 & 0 & 1 & 0 & 1 & 0 & 1 \\ 1 & \sqrt{2}/2 & \sqrt{2}/2 & 0 & 1 & -\sqrt{2}/2 & \sqrt{2}/2 & -1 \\ 1 & 0 & 1 & -1 & 0 & 0 & -1 & 1 \\ 1 & -\sqrt{2}/2 & \sqrt{2}/2 & 0 & -1 & \sqrt{2}/2 & \sqrt{2}/2 & -1 \\ 1 & -1 & 0 & 1 & 0 & -1 & 0 & 1 \\ 1 & -\sqrt{2}/2 & -\sqrt{2}/2 & 0 & 1 & \sqrt{2}/2 & -\sqrt{2}/2 & -1 \\ 1 & 0 & -1 & -1 & 0 & 0 & 1 & 1 \\ 1 & \sqrt{2}/2 & -\sqrt{2}/2 & 0 & -1 & -\sqrt{2}/2 & -\sqrt{2}/2 & -1 \end{bmatrix}$$

Now take a vector in $\mathbb{R}^8$, let $\mathbf{u} = \begin{bmatrix} -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 1 \\ 0 \end{bmatrix}$. Now $\mathbf{u}$ can be written as a linear combination of the columns of F and the required weights can be computed

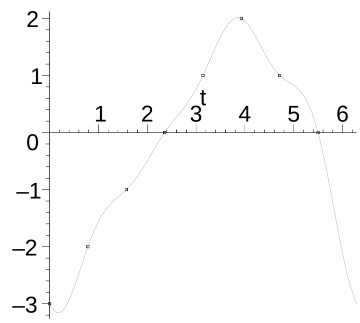
$$F^{-1}\mathbf{u} = \begin{bmatrix} -.2500000000 \\ -1.707106781 \\ -1.207106781 \\ -.50000000000 \\ 0 \\ -.2928932190 \\ -.2071067810 \\ -.2500000000 \end{bmatrix}$$

One way of interpreting this result is pretty trivial. We have a vector in $\mathbb{R}^8$ and a basis of $\mathbb{R}^8$ and we merely wrote the vector as a linear combination of the basis vectors. But there is a deeper interpretation. The basis vectors represent sine and cosine waves of different frequencies. If we combine these continuous functions using the weights that we computed we get

$$f(t) = -.25 - 1.707\cos(t) - 1.207\sin(t) - .5\cos(2t)$$

$$-.292\cos(3t) - .207\sin(3t) - .25\cos(4t)$$

This continuous function $f(t)$ gives us a waveform which, if sampled at $k\pi/4$ for $k = 0, 1, \ldots, 7$, would give us vector $\mathbf{u}$.
Plotting the discrete vector $\mathbf{u}$ and $f(t)$ we get the following



We chose an arbitrary vector in $\mathbb{R}^8$. That vector can be seen as being composed of components at different frequencies. When we express that vector in terms of the Fourier basis we get the weights of the components at the various frequencies.

Exercises

1. Let $\mathbf{v} = \begin{bmatrix} 1 & 2 & 4 & 4 \end{bmatrix}^T$. Write $\mathbf{v}$ as a linear combination of the Fourier basis for $\mathbb{R}^4$.

2. (a) Let $\mathbf{v} = \begin{bmatrix} 1 & 1 & -1 & -1 \end{bmatrix}^T$. Find the coordinates of $\mathbf{v}$ relative to the Fourier basis for $\mathbb{R}^4$.

   (b) Let $\mathbf{v} = \begin{bmatrix} 1 & 1 & 1 & 1 & -1 & -1 & -1 & -1 \end{bmatrix}^T$. Find the coordinates of $\mathbf{v}$ relative to the Fourier basis for $\mathbb{R}^8$.

3. (a) Sample $\cos^2 t$ at $t = 0, \pi/2, \pi, 3\pi/2$ and express the resulting vector in terms of the Fourier basis for $\mathbb{R}^4$.

   (b) Sample $\cos^2 t$ at $t = 0, \pi/4, \pi/2, 3\pi/4, \pi, 5\pi/4, 3\pi/2, 7\pi/4$ and express the resulting vector in terms of the Fourier basis for $\mathbb{R}^8$.

4. Let $\mathcal{F}$ be the discrete Fourier basis for $\mathbb{R}^4$ and let $\mathcal{H}$ be the Haar basis for $\mathbb{R}^4$. What is $P_{\mathcal{F}\leftarrow\mathcal{H}}$?

5. If a set of continuous functions is orthogonal it is not necessarily true that if you sample these functions you will obtain a set of orthogonal vectors. Show that if you sample the Legendre polynomials at equally spaced intervals over [-1, 1] the resulting vectors won't be orthogonal in the standard inner product.

Using MAPLE

Example 1.

In this example we will use the lynx population data that we saw in Chapter 1. We will convert this to a Fourier basis and plot the amplitudes of the components at each frequency.

```
>u:=readdata(`lynx.dat`):
```

We now have the data entered into **Maple** as a list called $\mathbf{u}$. There are two problems at this point. First, we will be using a procedure that is built into **Maple** for converting $\mathbf{u}$ to a Fourier basis and this procedure only works on vectors whose length is a power of 2. This is a common problem and one standard solution is to use zero-padding to increase the length of our data. We will add zeroes to vector $\mathbf{u}$ to bring it up to length 128. How many zeroes do we need to add?

```
>128-nops(u);
                  14
```

The `nops` command gives us the number of operands in `u`, which in this case is just the number of entries in `u`. So we have to add 14 zeroes to our list.

```
>u:=[op(u),0$14];
```

The command `op(u)` removes the brackets from list `u`, we then add 14 zeroes and place the result inside a new set of brackets.

The other problem is that the **Maple** procedure we will use only operates on **arrays**, not on **lists**. So we will convert `u` to an array and we will need another array consisting only of zeroes.

```
>xr:=convert(u,array):
>xi:=convert([0$128],array):
```

The procedure we will use in **Maple** actually operates on complex data. The real and imaginary parts are placed in separate arrays which we've called `xr` and `xi`. In this case our data is real so all the imaginary terms are zeroes. We now use the **Maple** procedure:

```
>readlib(FFT):
>FFT(7,xr,xi):
```

The `readlib` command loads the required procedure into **Maple** and the `FFT` command converts our data to a Fourier basis. The first parameter i n the `FFT` command gives the length of the data as the corresponding power of 2. In this case we have $2^7 = 128$, so the parameter is 7. When we use the `FFT` command we get the coefficients of the cosines in the first array, `xr`, and the coefficients of the sines in the second array, `xi`. The net amplitude at each frequency can then be computed and plotted.

```
>data:=[seq( [i-1, sqrt(xr[i]^2+xi[i]^2)],i=1..128)]:
>plot(data,i=0..64);
```

This gives **Figure 5.39**

If you look at this plot you see a peak at frequency 0. This is just the average value of our data (actually it is the average scaled by 128). More important is the second peak which occurs at a value

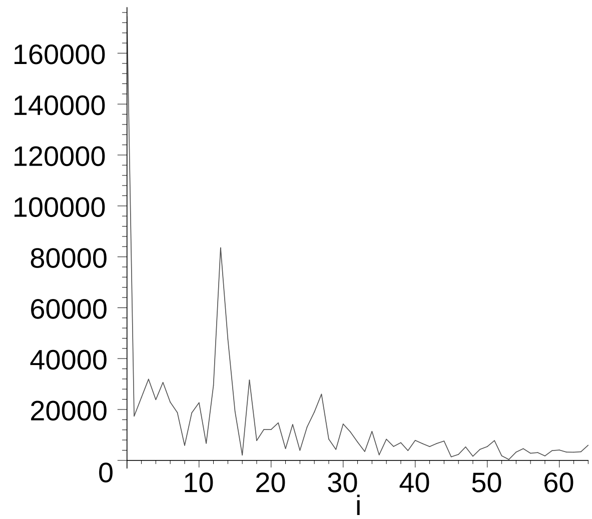


Figure 5.39: Frequency amplitudes for `lynx.dat`.

of 13 on the frequency axis. The lowest non-zero frequency in our Fourier basis would be 1 cycle/128 years. The frequency where this peak occurs is therefore 13 cycles/128 years or

$$\frac{128 \text{ years}}{13 \text{ cycles}} \approx 9.8 \text{ years/cycle}$$

In other words, this peak tells us that there is a cyclic pattern in our data that repeats roughly every 10 years.

There is a smaller peak at 27 which would correspond to a less prominent cycle which repeats every $\frac{128}{27} \approx 4.7$ years.

Example 2.

We will repeat Example 1 will some artificial data.

Suppose we have the function $f(t) = \sin(t) + .5\cos(4t) + .2\sin(9t)$, we will generate some discrete data by sampling this function 64 times on the interval $[0, 2\pi]$. We will then repeat the steps from example 1.

```
>f:=t->sin(t)+.5*cos(4*t)+.2*sin(9*t);
>u:=[seq(evalf(f(i*2*Pi/64)),i=1..64)];
>plots[listplot](u,style=point);
```

The last line above plots our data and gives us **Figure 5.40**

We now continue as before

```
>xr:=convert(u,array):
>xi:=convert([0$64],array):
>FFT(6,xr,xi):
>data:=[seq( [i-1, sqrt(xr[i]^2+xi[i]^2)],i=1..64)]:
>plot(data,x=0..32);
```

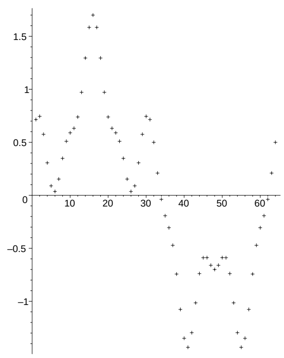


Figure 5.40: Our artificial data.

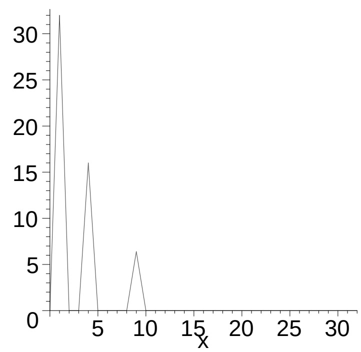


Figure 5.41: Frequency content of data.

We get **Figure 5.41**.

The plot shows the (relative) amplitude of the frequency content.

We'll do one more example of the same type. Let $g(t) = \sin(20t + 1.3\cos(4t))$. We will sample this 128 times on $[0, 2\pi]$.

```
>f:=t->sin(20*t+1.3*cos(4*t));
>u:=[seq(evalf(f(i*2*Pi/128)),i=1..128)];
>xr:=convert(u,array):
>xi:=convert([0$128],array):
>FFT(7,xr,xi):
>data:=[seq( [i-1, sqrt(xr[i]^2+xi[i]^2)],i=1..128)]:
>plot(data,x=0..64);
```

We get the plot **Figure 5.42**.

This shows that our vector had frequency content at many different values centered at 20 at spreading out from there. This is typical of **frequency modulation** as used in FM radio transmission.

Example 3.

In this example we will define a function composed of two sine waves at different frequencies with random numbers added. The addition of random numbers simulates the effect of noise on the data (measurement errors, outside interference, etc.). We use the **Maple** procedure `normald` to generate the random numbers. This example will illustrate two main uses of the Discrete Fourier Transform: it can be used to analyze the frequency content of a digitized signal, and (2) it can be used to modify the frequency content. (In this case to remove noise from the signal.)

```
>with(stats[random],normald):
>f:=x->evalf(sin(11*x)+.8*sin(5*x))+normald();
>xr:=[seq(f(i*Pi/64),i=1..128)]:
>plot([seq([i,xr[i]],i=1..128)]);
```

We get **Figure 5.43**

The addition of the noise has obscured the wave pattern of the sines.

Next we will look at the frequency content of the signal following the same steps as our previous example.

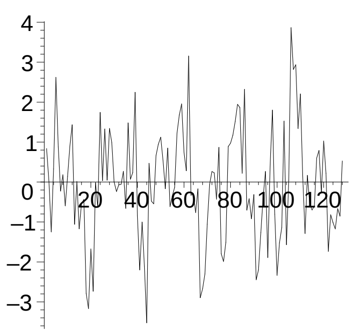


Figure 5.42: Noisy data.

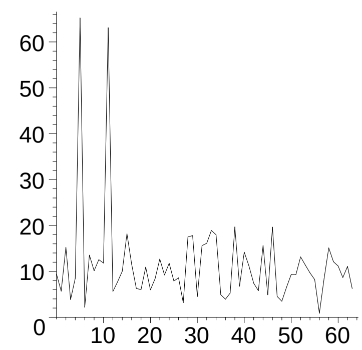


Figure 5.43: Frequency content of noisy data.

```
>xr:=convert(xr,array):
>yr:=convert([0$128],array):
>FFT(7,xr,yr):
>amps:=[seq(sqrt(xr[i]^2+yr[i]^2),i=1..64)]:
>plot([seq([i-1,amps[i]],i=1..64)]);
```

This **Figure 5.44**.

The two large spikes in this plot reveal the presence of the two sine components in our signal.

This is actually quite amazing. In summary here's what happened:

- We had a set of data values. By printing out this data and looking at the specific numerical values we would learn nothing nothing about the data. We would just see a list of apparently random numbers. Plotting the data is more helpful that looking at the raw numbers but still doesn't reveal much about the data.
- We then converted the data vector to a new basis, the discrete Fourier basis.
- By looking at the coordinates in our new basis we are able to discern a pattern in our data. We see that the original data consisted of primarily two periodic cycles. We can see the frequency of these cycles.

Since the noise seems to have turned up as lower amplitude data (all the frequencies apart from the two main components have amplitudes less than 20) we can remove it and just keep the peaks as follows:

```
>for i to 128 do
  if sqrt(xr[i]^2+yr[i]^2)<20 then xr[i]:=0; yr[i]:=0 fi
  od:
```

This line loops through the transformed data. If the amplitude at any frequency is less than 20 it makes the corresponding entries in `xr` and `yr` equal to 0. The intention is to remove the noise but keep the periodic data.

Next convert back to the standard basis. The **Maple** command for this is `iFFT`.

```
>iFFT(7,xr,yr);
>plot([seq([i,xr[i]],i=1..128)]);
```

We get **Figure 5.45**

The smoothness of the plot is the result of the noise having been removed. (Compare this plot with the plot of $.8\sin(5t) + .8 * \sin(11t)$.)

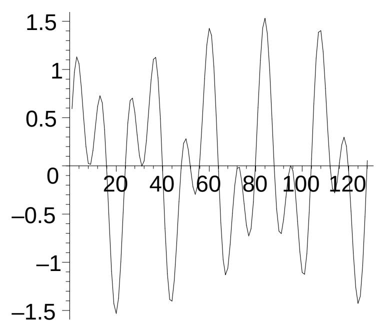


Figure 5.44: Reconstructed signal.

Chapter 6

Symmetric Matrices

So far in this course we have looked at two important sources for bases of $\mathbb{R}^n$. We saw that in certain situations it is convenient to construct a basis of eigenvectors of some matrix A. In other situations it is convenient to have a basis of orthogonal (or orthonormal) vectors. In this chapter these two streams will merge in the study of symmetric matrices. Symmetric matrices are the most important class of matrices that arise in applications.

6.1 Symmetric Matrices and Diagonalization

Definition 17 *A square matrix A is said to be* ***symmetric*** *if $A^T = A$. This means that A is symmetric if and only if $a_{ij} = a_{ji}$ for each entry in A.*

The following are examples of symmetric matrices (the third example assumes that you are familiar with the notation for partitioned matrices). Notice how each entry above the main diagonal is mirrored by an entry below the main diagonal:

$$\begin{bmatrix} 1 & 4 \\ 4 & 0 \end{bmatrix} \qquad \begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & 0 \\ -1 & 0 & 2 \end{bmatrix} \qquad \begin{bmatrix} 0 & A \\ A^T & 0 \end{bmatrix}$$

There is one particularly important property of symmetric matrices that will be illustrated by the next example.

Example 6.1.1

L et $A = \begin{bmatrix} 3 & 2 & 0 \\ 2 & 4 & 2 \\ 0 & 2 & 5 \end{bmatrix}$. The characteristic polynomial of this symmetric matrix is

$$-\lambda^3 + 12\lambda^2 - 39\lambda + 28 = -(\lambda - 7)(\lambda - 4)(\lambda - 1)$$

The eigenvalues are therefore 7, 4, and 1.
If we find a basis for each eigenspace we get

$$\lambda_1 = 7,\ \mathbf{v}_1 = \begin{bmatrix} 1 \\ 2 \\ 2 \end{bmatrix} \qquad \lambda_2 = 4,\ \mathbf{v}_2 = \begin{bmatrix} 2 \\ 1 \\ -2 \end{bmatrix} \qquad \lambda_3 = 1,\ \mathbf{v}_3 = \begin{bmatrix} 2 \\ -2 \\ 1 \end{bmatrix}$$

It is easy to verify that these three eigenvectors are mutually orthogonal. If we normalize these eigenvectors we obtain an orthogonal matrix, P, that diagonalizes A.

$$P = \begin{bmatrix} 1/3 & 2/3 & 2/3 \\ 2/3 & 1/3 & -2/3 \\ 2/3 & -2/3 & 1/3 \end{bmatrix}$$

Since P is orthogonal we have $P^{-1} = P^T$ and so

$$A = PDP^T = P \begin{bmatrix} 7 & 0 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{bmatrix} P^T$$

Definition 18 *A square matrix is said to be* ***orthogonally diagonalizable*** *if there is an orthogonal matrix P and a diagonal matrix D such that $A = PDP^T$.*

As we will soon see it turns out that *every* symmetric matrix can be orthogonally diagonalized. In order for this to happen it is necessary for a symmetric matrix to have orthogonal eigenvectors. The next theorem deals with this.

Theorem 6.1 *If A is a symmetric matrix then any two eigenvectors from different eigenspaces are orthogonal.*

Proof. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be eigenvectors corresponding to distinct eigenvalues λ_1 and λ_2 respectively. We want to show that $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$.

$$\begin{aligned} \lambda_1\mathbf{v}_1 \cdot \mathbf{v}_2 &= (\lambda_1\mathbf{v}_1)^T\mathbf{v}_2 & \\ &= (A\mathbf{v}_1)^T\mathbf{v}_2 & \text{Since } \mathbf{v}_1 \text{ is an eigenvector} \\ &= \mathbf{v}_1^T A^T\mathbf{v}_2 & \\ &= \mathbf{v}_1^T A\mathbf{v}_2 & \text{Since } A \text{ is symmetric} \\ &= \mathbf{v}_1^T(\lambda_2\mathbf{v}_2) & \text{Since } \mathbf{v}_2 \text{ is an eigenvector} \\ &= \lambda_2\mathbf{v}_1 \cdot \mathbf{v}_2 & \end{aligned}$$

We then get $\lambda_1\mathbf{v}_1 \cdot \mathbf{v}_2 = \lambda_2\mathbf{v}_1 \cdot \mathbf{v}_2$, and so $(\lambda_1 - \lambda_2)\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. But $\lambda_1 - \lambda_2 \neq 0$ since the eigenvalues are distinct, and so we must have $\mathbf{v}_1 \cdot \mathbf{v}_2 = 0$. ■

The above result means that any two eigenspaces of a symmetric matrix are orthogonal.

Theorem 6.2 *If A is symmetric then A has only real eigenvalues.*

Proof. Suppose A is symmetric and λ is an eigenvalue (possibly complex) with corresponding eigenvector $\mathbf{v}$ (possibly complex). We then have $A\mathbf{v} = \lambda\mathbf{v}$ and $A\overline{\mathbf{v}} = \overline{\lambda}\overline{\mathbf{v}}$. Therefore,

$$\begin{aligned} \lambda\overline{\mathbf{v}} \cdot \mathbf{v} &= \overline{\mathbf{v}}^T A\mathbf{v} \\ &= (A\mathbf{v})^T\overline{\mathbf{v}} \\ &= \mathbf{v}^T A^T\overline{\mathbf{v}} \\ &= \mathbf{v}^T A\overline{\mathbf{v}} \\ &= \overline{\lambda}\overline{\mathbf{v}} \cdot \mathbf{v} \end{aligned}$$

We then have $\lambda\|\mathbf{v}\|^2 = \overline{\lambda}\|\mathbf{v}\|^2$, and so $(\lambda - \overline{\lambda})\|\mathbf{v}\|^2 = 0$. Since $\|\mathbf{v}\| \neq 0$ (because $\mathbf{v}$ is an eigenvector) we must have $\lambda - \overline{\lambda} = 0$ and so λ must be real. ■

Theorem 6.3 *A square matrix A is orthogonally diagonalizable if and only if A is symmetric.*

Proof. Half of the proof is easy. Suppose A is orthogonally diagonalizable, then $A = PDP^T$. We then have $A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$ and so A is symmetric.

What about the other half of the proof? We have to show that if A is symmetric then we can get an orthogonal eigenbasis. The orthogonality is not a problem.The Gram-Schmidt procedure allows us to find an orthonormal basis for each eigenspace. Furthermore **Theorem 6.1** tells us that the eigenspaces of A are mutually orthogonal so the orthonormal bases for each eigenspace can be combined into one set of orthonormal vectors. But how do we know that A is not defective? That is, how do we know that we will get enough eigenvectors to form a basis? If A has distinct eigenvalues this is not a problem. But if A has a repeated eigenvalue how do we know that the eigenspace has the maximum possible dimension? The proof is fairly complicated and will not be given here but those interested can look in Appendix D where a proof is presented. ■

The following key theorem[1] summarizes the important properties of symmetric matrices that we have mentioned.

Theorem 6.4 (The Spectral Theorem for Symmetric Matrices) *If A is a symmetric $n \times n$ matrix then*

1. *A has n real eigenvalues counting multiplicities.*
2. *The dimension of each eigenspace is equal to the multiplicity of the corresponding eigenvalue.*
3. *The eigenspaces are mutually orthogonal.*
4. *A is orthogonally diagonalizable.*

Spectral Decomposition

The factorization of a symmetric matrix $A = PDP^T$ is called the spectral decomposition of the matrix. An alternate way of writing this is

$$\begin{aligned} A = PDP^T &= \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \dots & \mathbf{v}_n \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & & \vdots & \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1\mathbf{v}_1 & \lambda_2\mathbf{v}_2 & \dots & \lambda_n\mathbf{v}_n \end{bmatrix} \begin{bmatrix} \mathbf{v}_1^T \\ \mathbf{v}_2^T \\ \vdots \\ \mathbf{v}_n^T \end{bmatrix} \\ &= \lambda_1\mathbf{v}_1\mathbf{v}_1^T + \lambda_2\mathbf{v}_2\mathbf{v}_2^T + \dots + \lambda_n\mathbf{v}_n\mathbf{v}_n^T \end{aligned}$$

Notice when you write the spectral decomposition of A as

$$A = \lambda_1\mathbf{v}_1\mathbf{v}_1^T + \lambda_2\mathbf{v}_2\mathbf{v}_2^T + \dots + \lambda_n\mathbf{v}_n\mathbf{v}_n^T$$

you are breaking A up into a sum of rank 1 matrices. Each of these rank 1 matrices has the form $\lambda_i\mathbf{v}_i\mathbf{v}_i^T$ and you should recognize this as a projection matrix and a scalar. Each term of this

[1]The set of eigenvalues of a matrix A is sometimes called the *spectrum* of A.This is why the theorem is called a *spectral theorem*.

type orthogonally projects a vector into an eigenspace and then scales the projected vector by the corresponding eigenvalue.

Example 6.1.2

L et $A = \begin{bmatrix} 7/4 & -5/4 \\ -5/4 & 7/4 \end{bmatrix}$. This matrix has eigenvalues $\lambda_1 = 1/2$ and $\lambda_2 = 3$ with corresponding (normalized) eigenvectors $\mathbf{v}_1 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}$ and $\mathbf{v}_2 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix}$. Notice that the eigenspaces are orthogonal.
The spectral decomposition of A can be written as

$$A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T = \frac{1}{2} \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} + 3 \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} \end{bmatrix}$$

Expanding the right hand side we get

$$= \frac{1}{2} \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{bmatrix} + 3 \begin{bmatrix} 1/2 & -1/2 \\ -1/2 & 1/2 \end{bmatrix} = \begin{bmatrix} 7/4 & -5/4 \\ -5/4 & 7/4 \end{bmatrix}$$

Suppose we let $\mathbf{v} = \begin{bmatrix} 2 \\ 1 \end{bmatrix}$. Then $A\mathbf{v} = \begin{bmatrix} 9/4 \\ -3/4 \end{bmatrix}$. This multiplication by A can be seen as projections onto the eigenspaces combined with scaling by the corresponding eigenvalues. **Figure 6.1** shows vector $\mathbf{v}$ and the projections of $\mathbf{v}$ onto the two orthogonal eigenspaces. When $\mathbf{v}$ is multiplied by A each of the projections is scaled by the corresponding eigenvalue. When these scaled vectors are added we get $A\mathbf{v}$. This is shown in **Figure 6.2**.

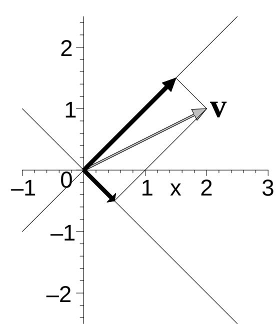


Figure 6.1: $\mathbf{v}$ and the eigenspaces of A.

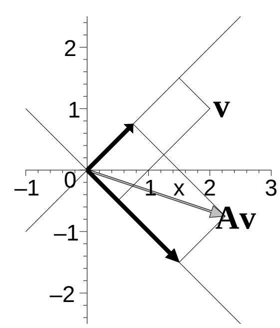


Figure 6.2: Av

Example 6.1.3

Orthogonally diagonalize the following matrix

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

It is easy to find the eigenvalues of 3 and 0 (multiplicity 2). Since A is not defective the eigenvalue 0 must correspond to a 2 dimensional eigenspace.

For $\lambda = 3$, proceeding in the usual way we have

$$\left[\begin{array}{rrr|r} -2 & 1 & 1 & 0 \\ 1 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right] \sim \left[\begin{array}{rrr|r} 1 & 0 & -1 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and this gives a basis of $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ for the corresponding eigenspace.

For $\lambda = 0$ we have

$$\left[\begin{array}{rrr|r} 1 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right]$$

and this gives a basis of $\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix} \right\}$.

It is also easy to see that the eigenspaces are indeed orthogonal. For $\lambda = 3$ the eigenspace is the line $\mathbf{x} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$. For $\lambda = 0$ the eigenspace is the plane $x_1 + x_2 + x_3 = 0$ with normal vector $\begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Matrix A could therefore be diagonalized by the matrix

$$\begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \\ 1 & 0 & -1 \end{bmatrix}$$

But this matrix does not have orthogonal columns. This is because one of the eigenspaces is 2 dimensional and the basis we chose for that eigenspace was not orthogonal. If we want an orthogonal basis for that eigenspace we can use the Gram-Schmidt procedure and get

$$\left\{ \begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ -2 \end{bmatrix} \right\}$$

as an orthogonal basis of the 2 dimensional eigenspace.

Now if we normalize our new basis vectors and use them as the columns of a matrix we get

$$P = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{6} \\ 1/\sqrt{3} & 0 & -2/\sqrt{6} \end{bmatrix}$$

and

$$P^T A P = \begin{bmatrix} 3 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Note that in this example if we call the one dimensional eigenspace W then the other eigenspace would be $W^{\perp}$. If we let $P = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \mathbf{v}_3 \end{bmatrix}$ then the matrix

$$\mathbf{v}_1\mathbf{v}_1^T = \begin{bmatrix} 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \\ 1/3 & 1/3 & 1/3 \end{bmatrix}$$

projects vectors orthogonally onto W. The matrix

$$\mathbf{v}_2\mathbf{v}_2^T + \mathbf{v}_3\mathbf{v}_3^T = \begin{bmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ -1/3 & -1/3 & 2/3 \end{bmatrix}$$

projects vectors orthogonally onto $W^{\perp}$.
You should also notice that $\mathbf{v}_1\mathbf{v}_1^T$ has rank 1. This means that the dimension of its column space is 1, and this corresponds to the fact that it projects vectors onto a line. On the other hand $\mathbf{v}_2\mathbf{v}_2^T + \mathbf{v}_3\mathbf{v}_3^T$ has rank 2 and this corresponds to the fact that it projects vectors onto a plane, a 2 dimensional eigenspace of A.

Example 6.1.4

L et $A = \begin{bmatrix} 0 & 2 \\ 2 & 3 \end{bmatrix}$. How does the unit circle get transformed when it is multiplied by A?
Figure 6.3 shows the unit circle before and after multiplication by A along with the eigenspaces of A. Now if we compute the spectral decomposition of A we get

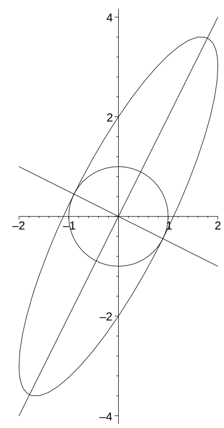


Figure 6.3: The unit circle transformed by A.

$$A = PDP^T = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{5} & 2/\sqrt{5} \\ -2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix}$$

Now we can look at the effect of multiplication by A as the combination of three factors: first we multiply by P^T, then by D, and finally by P. Multiplication by P^T rotates the unit circle clockwise by $\arccos(1/\sqrt{5})$ radians. P^T rotates the unit circle so that the orthogonal eigenspaces are rotated onto the axes. Next, multiplication by D scales the circle along the axes by the eigenvalues. In this case the circle is stretched by a factor of 4 along the x axis and flipped around the y axis resulting in an ellipse. Finally, multiplication by P rotates the ellipse so that the axes of the ellipse are now oriented along the eigenspaces.

If you think about this example it should be clear that if you multiply the unit circle by any 2×2 symmetric matrix then the same analysis can be applied and the result will always be an ellipse. The axes of the ellipse will be the eigenspaces of the matrix and the dimensions of the ellipse will be determined by the eigenvalues. (If one of the eigenvalues is zero, the result will be a degenerate ellipse. That is, a line segment.)

Skew-symmetric Matrices

Definition 19 *A matrix A is **skew-symmetric** if $A^T = -A$.*

An example of a skew-symmetric matrix is $\begin{bmatrix} 0 & 2 \\ -2 & 0 \end{bmatrix}$. A skew-symmetric matrix must have zeros on the diagonal and for the off-diagonal entries $a_{ij} = -a_{ji}$.

Remember that for a symmetric matrix all the eigenvalues must be real. For a skew-symmetric matrix we have the following result:

Theorem 6.5 *If A is skew-symmetric then the only possible real eigenvalue is 0.*

Proof. Let A be skew-symmetric and suppose that $A\mathbf{x} = \lambda\mathbf{x}$ for (possibly complex) λ and $\mathbf{x}$. Then

$$\begin{aligned}
\lambda\|\mathbf{x}\|^2 &= \lambda\mathbf{x}^T\overline{\mathbf{x}} \\
&= (A\mathbf{x})^T\overline{\mathbf{x}} \\
&= \mathbf{x}^T A^T\overline{\mathbf{x}} \\
&= \mathbf{x}^T(-A\overline{\mathbf{x}}) \\
&= -\mathbf{x}^T A\overline{\mathbf{x}} \\
&= -\overline{\lambda}\|\mathbf{x}\|^2
\end{aligned}$$

Since we then have $\lambda\|\mathbf{x}\|^2 = -\overline{\lambda}\|\mathbf{x}\|^2$ and $\|\mathbf{x}\| \neq 0$ we must have $\lambda = -\overline{\lambda}$. If λ is real this means that $\lambda = 0$. ■

So if a (real) matrix is symmetric all the eigenvalues lie on the real axis in the complex plane. If a matrix is skew-symmetric then all the eigenvalues lie on the imaginary axis in the complex plane.

Example 6.1.5

L et $\mathbf{u} = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$. Let $\mathbf{v}$ be any vector in $\mathbb{R}^3$. Then

$$\mathbf{u} \times \mathbf{v} = \begin{bmatrix} 0 & -c & b \\ c & 0 & -a \\ -b & a & 0 \end{bmatrix} \mathbf{v}$$

So the cross product of vectors in $\mathbb{R}^3$ can be looked at as a linear transformation corresponding to multiplication by a skew-symmetric matrix.

The characteristic polynomial of this skew-symmetric matrix would be

$$\lambda\left(\lambda^2 + a^2 + b^2 + c^2\right)$$

The eigenvalues are therefore 0 and $i\sqrt{a^2+b^2+c^2}$.
If you found the matrix of this transformation relative to another basis would the matrix of the transformation still be skew-symmetric?

Exercises

1. Orthogonally diagonalize the following symmetric matrices.

$$(a)\begin{bmatrix}1 & 3\\ 3 & 1\end{bmatrix}\qquad (b)\begin{bmatrix}0 & 2\\ 2 & 0\end{bmatrix}\qquad (c)\begin{bmatrix}16 & -4\\ -4 & 1\end{bmatrix}$$

$$(d)\begin{bmatrix}0 & 0 & 1\\ 0 & 1 & 0\\ 1 & 0 & 0\end{bmatrix}\qquad (e)\begin{bmatrix}1 & 1 & 3\\ 1 & 3 & 1\\ 3 & 1 & 1\end{bmatrix}\qquad (f)\begin{bmatrix}0 & 4 & 0\\ 4 & 0 & 3\\ 0 & 3 & 0\end{bmatrix}$$

2. Orthogonally diagonalize the following matrices:

$$(a)\begin{bmatrix}0 & 0 & 0 & 1\\ 0 & 1 & 0 & 0\\ 0 & 0 & 1 & 0\\ 1 & 0 & 0 & 0\end{bmatrix}\qquad (b)\begin{bmatrix}0 & 0 & 0 & 0 & 1\\ 0 & 0 & 0 & 1 & 0\\ 0 & 0 & 1 & 0 & 0\\ 0 & 1 & 0 & 0 & 0\\ 1 & 0 & 0 & 0 & 0\end{bmatrix}$$

3. Given that the matrix $A = \begin{bmatrix}1 & -1 & -1 & -1\\ -1 & 1 & -1 & -1\\ -1 & -1 & 1 & -1\\ -1 & -1 & -1 & 1\end{bmatrix}$ has eigenvalues 2 and -2 find an orthogonal matrix P such that P^TAP is diagonal. Find the matrices which orthogonally project vectors into each eigenspace.

4. Find the spectral decomposition of $\begin{bmatrix}a & b\\ b & a\end{bmatrix}$.

5. Find the spectral decomposition of $\begin{bmatrix}0 & a & b\\ a & 0 & 0\\ b & 0 & 0\end{bmatrix}$.

6. (a) Suppose P is a matrix that orthogonally diagonalizes matrix A, show that P also orthogonally diagonalizes matrix $A + kI$

   (b) Use the above result and the results from **Example 6.1.2** to orthogonally diagonalize

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

7. Let

$$\mathbf{v}_1 = \frac{1}{\sqrt{3}}\begin{bmatrix}1\\ 1\\ -1\end{bmatrix} \text{ and } \mathbf{v}_2 = \frac{1}{\sqrt{6}}\begin{bmatrix}1\\ 1\\ 2\end{bmatrix}$$

   Give a diagonal matrix that is similar to the following symmetric matrices:

   (a) $\mathbf{v}_1\mathbf{v}_1^T$

   (b) $\mathbf{v}_1\mathbf{v}_1^T + \mathbf{v}_2\mathbf{v}_2^T$

   (c) $5\mathbf{v}_1\mathbf{v}_1^T + 9\mathbf{v}_2\mathbf{v}_2^T$

8. Let $A = \begin{bmatrix} 0 & -2 & 2 \\ 2 & 0 & -1 \\ -2 & 1 & 0 \end{bmatrix}$. A is a skew-symmetric matrix. Find an orthogonal matrix P such that

$$P^T AP = \begin{bmatrix} 0 & 0 & 0 \\ 0 & a & -b \\ 0 & b & a \end{bmatrix}$$

for some real numbers a and b.

9. Given the spectral decomposition $A = \lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \cdots + \lambda_n \mathbf{v}_n \mathbf{v}_n^T$
   (a) Express A^2 in terms of the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$?
   (b) Express A^k in terms of the eigenvalues λ_i and eigenvectors $\mathbf{v}_i$?

10. Consider the linear transformation $T(\mathbf{x}) = A\mathbf{x}$ for each of the following matrices. Describe what effect the transformation has on the unit circle. In each case plot the unit circle before and after the the transformation along with any real eigenspaces of the transformation. Notice that one matrix is skew-symmetric, one is symmetric and invertible, and one is symmetric and not invertible.
    (a) $A = \begin{bmatrix} 0 & -4 \\ 4 & 0 \end{bmatrix}$
    (b) $A = \begin{bmatrix} 3 & 4 \\ 4 & -3 \end{bmatrix}$
    (c) $A = \begin{bmatrix} 1 & 4 \\ 4 & 16 \end{bmatrix}$

11. If A and B are symmetric $n \times n$ matrices show that $A + B$ is also symmetric.

12. If A and B are symmetric $n \times n$ matrices show by an example that AB is not necessarily symmetric.

13. Let $\mathbf{v}_1$ and $\mathbf{v}_2$ be orthogonal, unit vectors. Show that $\mathbf{v}_1 \mathbf{v}_1^T + \mathbf{v}_2 \mathbf{v}_2^T$ satisfies the two conditions of being an orthogonal projector. What subspace would this matrix project onto?

14. Suppose A is an $n \times n$ matrix. Show that $A^T A$ and AA^T are symmetric.

15. For each of the following statements give a proof to show that it is true or give a counter-example to show that it is false.
    (a) If A is orthogonal then A is orthogonally diagonalizable.
    (b) If A is orthogonally diagonalizable then A is orthogonal.
    (c) If A is orthogonally diagonalizable then so is $A + I$.

16. (a) If A is an $n \times n$ matrix show that $A + A^T$ is symmetric.
    (b) If A is an $n \times n$ matrix show that $A - A^T$ is skew-symmetric.
    (c) Use the last two results to show that any $n \times n$ matrix A can be written as the sum of a symmetric and a skew-symmetric matrix.

17. If A is a skew-symmetric $n \times n$ matrix and $\mathbf{v}$ is a vector in $\mathbb{R}^n$ show that $A\mathbf{v}$ is orthogonal to $\mathbf{v}$.

18. Prove that if A is skew-symmetric then A^2 is symmetric.

19. If A is skew-symmetric show that $I - A$ is invertible.

Using MAPLE

Example 1.

First we will illustrate using **Maple** to find the spectral decomposition of a symmetric matrix. We will begin with the matrix

$$A = \begin{bmatrix} 1 & 1 & 3 & 0 \\ -1 & 2 & 2 & 1 \\ 3 & 1 & 1 & 1 \\ 2 & -2 & 0 & 1 \end{bmatrix}$$

Now A is not a symmetric matrix but, as seen in the exercises, the matrix $A + A^T$ will be symmetric. We will let $B = A + A^T$ and find the spectral decomposition of B. The `evalf` command in the first line below means that all our computations will be carried out in floating-point form. Remember that this inevitably means some rounding error will creep into our computations.

```
>A:=evalf(<<1&-1&3&2>|<1&2&1&-2>|<3&2&1&0>|<0&1&1&1>>):  ### use floating point computations
>B:=A+A^%T;    ### B is symmetric
>ev,P:=Eigenvectors(B):
>P.B.P^%T;    ### B is diagonalized
>P^%T.P;      ### This will confirm that P is orthogonal
```

The last command returns the identity matrix (with some rounding error) which confirms that the columns of P form an orthonormal basis.

```
>P^%T.B.P;
>v[1]:=Column(P,1):v[2]:=Column(P,2):v[3]:=Column(P,3):v[4]:=Column(v,4):
>simplify( add( ev[i]*v[i].v[i]^%T, i=1..4));
```

The first line above diagonalizes B. It returns a diagonal matrix with the eigenvalues of B down the diagonal (again, with some rounding error). The second line defines the eigenvectors. The third line is the spectral decomposition

$$\lambda_1 \mathbf{v}_1 \mathbf{v}_1^T + \lambda_2 \mathbf{v}_2 \mathbf{v}_2^T + \lambda_3 \mathbf{v}_3 \mathbf{v}_3^T$$

expressed in **Maple** code and should return matrix B.

Example 2.

Let $A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 4 & 3 & 2 & 1 \end{bmatrix}$.

A is not a symmetric matrix but both $A^T A$ and AA^T are symmetric. We will use **Maple** to find the spectral decomposition of these two matrices.

```
>A:=<<1,4>|<2,3>|<3,2>|<4,1>>:    ### here we use exact values
>A1:=A.A^%T;    ### A1 is symmetric
>A2:=A^%T.A;    ### A2 is symmetric
>ev,V:=Eigenvectors(A1);
>V^%T.V; V^%T.A1.V;   ### ??? What does this mean

>ew,W:=Eigenvectors(A2);
>W^%T.W; W^%T.A2.W;   ### ??? What does this mean?
```

Notice that A^TA and AA^T have the same non-zero eigenvalues. A^TA is 3×3 and has 3 eigenvalues, and AA^T is 2×2 and has 2 eigenvalues. You should also be aware that **Maple** might not always return the eigenvalues and eigenvectors in the same order as indicated above.

Note that the computation V^TV shows that **Maple** has given us orthogonal eigenvectors, but not unit eigenvectos. So the columns of V must be nornmalized to get a matrix that orthogonally diagonalizes $A1$.

The computation W^TW is more complicated. This does not give a diagonal matrix. The coumns of W are not orthogonal. This is because the eigenspace corresponding to the eigenvalue 0 is 2 dimensional and **Maple** did not give an orthogonal basis for this eigenspace. We can use the Gram-Scmidt procedure, or the `QRDecomposition` command to fix this. In other words

```
>QRDecomposition(W,output='Q');
```

will return a matrix whose columns are orthonormal eigenvectors of $A2$.

Example 3.

In this example we will animate the relationship between the eigenspaces of a symmetric 2×2 matrix and the effect of multiplying the unit circle by this matrix. We will start with the matrix

$$A = \begin{bmatrix} 0 & k \\ k & 3 \end{bmatrix}$$

```
>with(plots):  ## we will need the animate command
>A:=<<0,k>|<k,3>>;
>v:=<cos(t),sin(t)>;
>tv:=A.v;
>ev,V:=Eigenvectors(A);
>v1:=Column(V,1):   ### an eigenvector
>v2:=Column(V,2):   ### the other eigenvector
>line1:=[ t*v1[1], t*v1[2], t=-5..5]: ### an eigenspace
>line2:=[ t*v2[1], t*v2[2], t=-5..5]: ### the other eigenspace
```

Now we can animate the results (remember the `animate` command must be loaded using `with(plots)`:

```
>animate(plot,[ { line1, line2, [tv[1],tv[2],t=-Pi..Pi] }],
                k=-3..3,view=[-5..5,-5..5],
                 scaling=constrained);
```

When you play the animation you see the ellipse and eigenspaces as k varies from -3 to 3. Notice that the eigenspaces remain the axes of the ellipse.

Example 4.

In this example we will define a matrix A. We will then define the symmetric matrix $A1 = 1/2(A+A^T)$ and the skew-symmetric matrix $A2 = 1/2(A+A^T)$. We will find the eigenvalues of each of these matrices and plot them in the complex plane. Since $A1$ is symmetric all of its eigenvalues should be on the real axis. All the eigenvalues of $A2$ should lie on the imaginary axis. The eigenvalues of A could be anywhere in the complex plane. We will start by defining A using the `randmatrix` command in **Maple** .

```
>with(plots):
>A:=RandomMatrix(20,generator=-1.0..1.0):
>A1:=evalm(.5*(A+transpose(A))):  ### A1 is symmetric
>A2:=evalm(.5*(A-transpose(A))):  ### A2 is skew-symmetric
>E:=convert(Eigenvalues(A),list):
>E1:=convert(Eigenvalues(A1),list):
>E2:=convert(Eigenvals(A2),list):
>p1:=complexplot(E,style=point,symbol=diamond):   ### in cojugate pairs
>p2:=complexplot(E1,style=point,symbol=circle):  ### on real axis
>p3:=complexplot(E2,style=point,symbol=box):   ### on imaginary axis
>display([p1,p2,p3]);
```

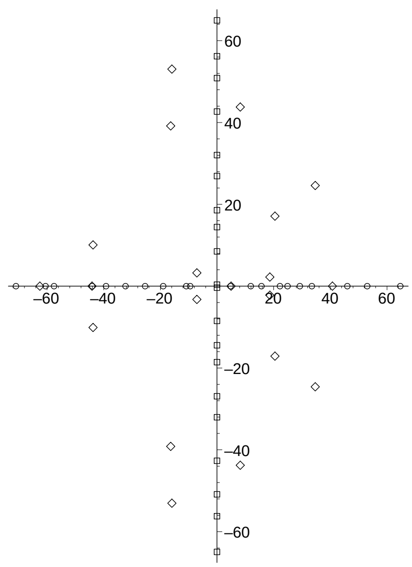


Figure 6.4: The eigenvalues in the complex plane.

Figure 6.4 shows a typical output of the above commands. The eigenvalues of $A2$ have ended up on the imaginary axis. The eigenvalues of $A1$ are on the real axis. The eigenvalues of A are spread out over the complex plane.

Here is a variation on this example. We begin as before.

```
>with(plots):
>A:=evalm(.2*randmatrix(20,20)):
>A1:=evalm(.5*(A+transpose(A))):
>A2:=evalm(.5*(A-transpose(A))):
```

Next we will consider matrices of the form $(1-t)A1+tA2$ for $0 \le t \le 1$. Notice that when $t=0$ this expression gives the symmetric matrix $A1$ and when $t=1$ this expression gives the skew-symmetric matrix $A2$.

```
>for i to 21 do
  t:=.05*(i-1):
  M:=evalm((1-t)*A1+t*A2):
  ev:=convert(Eigenvalues(M),list):
  p[i]:=complexplot(ev,style=point,symbol=diamond,color=black):
  od:
>display([seq( p[i],i=1..21)], insequence=true);
```

The animation begins with the eigenvalues of $A1$ (all on the real axis) and ends with the eigenvalues of $A2$ (all on the imaginary axis). The intermediate frames correspond to linear combinations of $A1$ and $A2$ that are neither symmetric nor skew-symmetric.

6.2 Quadratic Forms

Definition 20 *A* ***quadratic form*** *is an expression of the form* $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ *where* A *is a symmetric matrix and* $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$. *When expanded a quadratic form is a linear combination of terms of the form* $x_i x_j$ *(where* i *and* j *could be the same). The terms in a quadratic form where* $i \neq j$ *are called* ***cross product*** *terms.*

For example, let $A = \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix}$ and $\mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$ then

$$\begin{aligned} \mathbf{x}^T A\mathbf{x} &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 3 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \\ &= \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} x_1 + 2x_2 \\ 2x_1 + 3x_2 \end{bmatrix} \\ &= x_1(x_1 + 2x_2) + x_2(2x_1 + 3x_2) \\ &= x_1^2 + 4x_1x_2 + 3x_2^2 \end{aligned}$$

If you look carefully at the details of the above computation you should be able to see that the entries on the main diagonal of A correspond to the terms where one of the variables is being squared and the off diagonal entries combine to give the cross product term. In particular the coefficient of the cross-product term $x_i x_j$ is a combination of the terms a_{ij} and a_{ji}. So if you had the expression $5x_1 - 8x_1x_2 - 7x_2^2$ it could be written as $\mathbf{x}^T A\mathbf{x}$ where $A = \begin{bmatrix} 5 & -4 \\ -4 & -7 \end{bmatrix}$.

The same pattern applies to larger matrices also. So, for example, the symmetric matrix

$$A = \begin{bmatrix} 2 & 1 & -1 \\ 1 & 3 & 3 \\ -1 & 3 & 4 \end{bmatrix}$$

would generate the quadratic form $2x_1^2 + 3x_2^2 + 4x_3^2 + 2x_1x_2 - 2x_1x_3 + 6x_2x_3$.

Similarly, the quadratic form $x_1^2 + x_1x_2 - 3x_3x_4$ would be generated by the matrix

$$\begin{bmatrix} 1 & 1/2 & 0 & 0 \\ 1/2 & 0 & 0 & 0 \\ 0 & 0 & 0 & -3/2 \\ 0 & 0 & -3/2 & 0 \end{bmatrix}$$

The key property of quadratic forms that we will be using is that all cross product terms can be eliminated from a quadratic form by using a change of variables. Specifically, suppose that we have $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$. Since A is symmetric we have $A = PDP^T$ for an orthogonal matrix P and a diagonal matrix D. Therefore,

$$\begin{aligned} Q(\mathbf{x}) &= \mathbf{x}^T A\mathbf{x} \\ &= \mathbf{x}^T PDP^T\mathbf{x} \\ &= (P^T\mathbf{x})^T DP^T\mathbf{x} \\ &= \mathbf{y}^T D\mathbf{y} \qquad \text{where} \qquad \mathbf{y} = P^T\mathbf{x} \end{aligned}$$

Since D is diagonal, the quadratic form $\mathbf{y}^T D\mathbf{y}$ has no cross product terms, so the change of variable has resulted in a simpler, but equivalent, quadratic form. You should also notice that since P is orthogonal the change of variable equation can also be written as $\mathbf{x} = P\mathbf{y}$.

Example 6.2.6

Suppose $A = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$ so

$$\mathbf{x}^T A\mathbf{x} = 3x_1^2 + 2x_1x_2 + 3x_2^2$$

When we compute the spectral decomposition of A we get

$$A = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} 4 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix}$$

We therefore introduce a new coordinate system by letting

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} \sqrt{2}/2 & \sqrt{2}/2 \\ -\sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

We then have $Q(\mathbf{x}) = 4y_1^2 + 2y_2^2$.
Now suppose we want to plot the equation $3x_1^2 + 2x_1x_2 + 3x_2^2 = 8$. From the previous result, this is equivalent to plotting $4y_1^2 + 2y_2^2 = 8$, and this would be very easy to plot. You should recognize this as the equation of an ellipse with intercepts $y_1 = \pm\sqrt{2}$ and $y_2 = \pm 2$. But this would be the plot relative to the $\mathbf{y}$ coordinate system. What is the plot relative to the $\mathbf{x}$ coordinate system? If we look back at the relation $\mathbf{y} = P^T\mathbf{x}$ we see that the $\mathbf{y}$ coordinate system in this case is just the $\mathbf{x}$ coordinate system rotated by $\pi/4$ (just look at the entries in P) [2]. Another way of saying this this is that the y_1 and y_2 axes are just the orthogonal eigenspaces of the matrix A that generated the original quadratic form.
Figure 6.5 is the plot of this equation. The y_1 and y_2 axes are the orthogonal eigenspaces of A with bases $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} -1 \\ 1 \end{bmatrix}$ repectively. The axes of the ellipse are long these eigenspaces.
What are the points on the plot that are closest to the origin? If you look at the plot it should be clear that the points closest to the origin are the y_1 intercepts.
What are the coordinates of these points? This question is a bit ambiguous since there are two different coordinate systems being used. Let's look at each coordinate system. In the $\mathbf{y}$ system it should be easy to see from the equation that the closest points are $(\pm\sqrt{2}, 0)$. In the $\mathbf{x}$ system the coordinates are given by $\mathbf{x} = P\mathbf{y} = \begin{bmatrix} \sqrt{2}/2 & -\sqrt{2}/2 \\ \sqrt{2}/2 & \sqrt{2}/2 \end{bmatrix} \begin{bmatrix} \pm\sqrt{2} \\ 0 \end{bmatrix} = \pm \begin{bmatrix} 1 \\ 1 \end{bmatrix}$.
Similarly, the points on the ellipse that are furthest away from the origin would be $(0, \pm 2)$ in the $\mathbf{y}$ system and $\pm(\sqrt{2}, -\sqrt{2})$ in the $\mathbf{x}$ system.

[2]There is a common source of confusion here. The y_1 axis is in the direction $\begin{bmatrix} 1 \\ 0 \end{bmatrix}$ in the y coordinate system. In the x system this vector would be $P\begin{bmatrix} 1 \\ 0 \end{bmatrix}$. In other words, the y axes are the x axes rotated by P. But to convert from the x coordinates of a vector to the y coordinates we multiply by P^T. It's simplest just to remember that the y axes are just the eigenspaces of A.

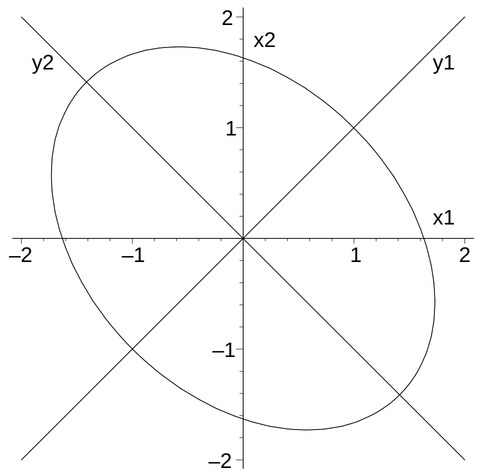


Figure 6.5: $3x_1^2 + 2x_1x_2 + 3x_2^2 = 8$

The following theorem summarizes the points made so far.

Theorem 6.6 (The Principal Axes Theorem) *Let A be a symmetric $n \times n$ matrix. There is an orthogonal change of variable $\mathbf{x} = P\mathbf{y}$ that transforms the quadratic form $\mathbf{x}^T A\mathbf{x}$ into a quadratic form $\mathbf{y}^T D\mathbf{y}$ with no cross-product terms.*

The eigenspaces of matrix A in the theorem are called the **principal axes** of the quadratic form if the eigenvalues are all distinct[3]. The columns of matrix P in the theorem are bases for these eigenspaces.

Example 6.2.7

For another example we'll plot the equation $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$. This could be written as $\mathbf{x}^T A\mathbf{x} = 16$ where

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix}$$

A has the characteristic polynomial

$$\lambda^2 + 4\lambda - 21 = (\lambda - 7)(\lambda + 3)$$

This gives us eigenvalues of -7 and 3. The eigenspace for $\lambda = -7$ has basis $\begin{bmatrix} 1/\sqrt{5} \\ 2/\sqrt{5} \end{bmatrix}$. The eigenspace corresponding to $\lambda = 3$ will be orthogonal to the first eigenspace so it will have basis $\begin{bmatrix} -2/\sqrt{5} \\ 1/\sqrt{5} \end{bmatrix}$. Giving us

$$P = \begin{bmatrix} 1/\sqrt{5} & -2/\sqrt{5} \\ 2/\sqrt{5} & 1/\sqrt{5} \end{bmatrix} \qquad \text{and} \qquad \mathbf{y}^T D\mathbf{y} = -7y_1^2 + 3y_2^2$$

[3]Suppose the eigenvalues were not distinct. For example, suppose some eigenvalue had multiplicity 2. The corresponding eigenspace would the be a plane, and there would be an infinite number of choices for an orthonormal basis of this eigenspace. There would therefore be no uniquely defined pair of axes lying in this plane.

The original equation is therefore equivalent to $-7y_1^2 + 3y_2^2 = 16$. You should recognize this as the graph of a hyperbola. There is no y_1 intercept and the y_2 intercepts are $\pm 4/\sqrt{3}$. The hyperbola is oriented along the eigenspaces of A. **Figure 6.6** shows a plot of the hyperbola and its principal axes.

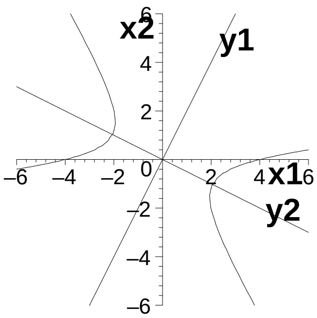


Figure 6.6: $x_1^2 - 8x_1x_2 - 5x_2^2 = 16$

Example 6.2.8

In this example we will find the points on the graph of

$$-x_1^2 + x_2^2 - x_3^2 + 10x_1x_3 = 1$$

that are closest to the origin.

The idea in this section is that a question like the above becomes much simpler to answer if we look at it from the point of view of the principal axes. The given quadratic form is generated by matrix

$$A = \begin{bmatrix} -1 & 0 & 5 \\ 0 & 1 & 0 \\ 5 & 0 & -1 \end{bmatrix}$$

which has characteristic polynomial $(1-\lambda)(\lambda-4)(\lambda+6)$ so we get eigenvalues of 1, 4, and -6 with corresponding eigenvectors $\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$.

So the original problem can be seen as the simpler problem of finding the points closest to the origin on the graph of $y_1^2 + 4y_2^2 - 6y_3^2 = 1$. The closest points would be the y_2 intercepts $(0, \pm 1/2, 0)$. In the original coordinate system we have $\mathbf{x} = P\mathbf{y}$ which gives

$$\begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 \\ \pm 1/2 \\ 0 \end{bmatrix} = \pm \begin{bmatrix} \sqrt{2}/4 \\ 0 \\ \sqrt{2}/4 \end{bmatrix}$$

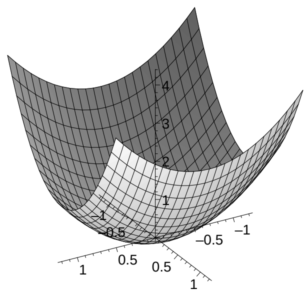


Figure 6.7: $Q(\mathbf{x}) = x_1^2 + 2x_2^2$

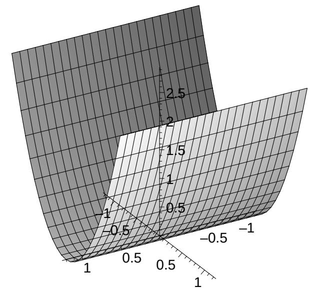


Figure 6.8: $Q(\mathbf{x}) = 2x_2^2$

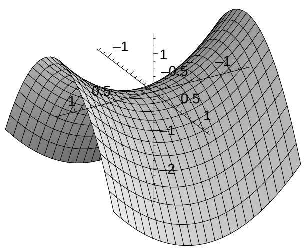


Figure 6.9: $Q(\mathbf{x}) = x_1^2 - 2x_2^2$

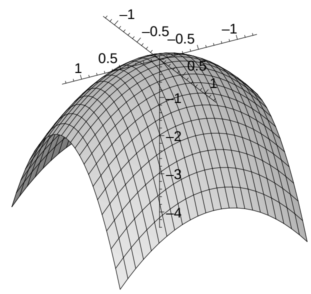


Figure 6.10: $Q(\mathbf{x}) = -x_1^2 - 2x_2^2$

Classification of Quadratic forms

A quadratic form, $\mathbf{x}^T A\mathbf{x}$, is typically classified by the range of values it may possibly assume for various values of $\mathbf{x}$. In **Figures 6.7 - 6.6.10** we see the plots of four related quadratic forms. The two horizontal axes are x_1 and x_2 and the vertical axis corresponds to the value of the quadratic form.

In **Figure 6.7** both terms of the quadratic form are positive. The corresponding surface is called a **paraboloid**. It should be clear that this quadratic form is never negative, and that $Q(\mathbf{x})$ takes on a minimum of 0 at the origin.

In **Figure 6.8** the quadratic form is again never negative. This surface is called a **parabolic cylinder**. The minimum value is again 0 but now there are infinitely many points where this minimum occurs. It is taken on at all points where $x_2 = 0$.

In **Figure 6.9** we have one positive coefficient and one negative. Here the quadratic form takes on both positive and negative values. The corresponding surface is called a **hyperbolic paraboloid** or more informally a **saddle curve**. In this case there is no maximum or minimum value of the quadratic form. The origin itself is called a **saddle point**.

In **Figure 6.10** both coefficients are negative. This again gives a paraboloid except that it opens downward rather that upward. This makes the origin a maximum.

Definition 21 *A quadratic form* $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ *is said to be:*

a. **positive definite** *if* $Q(\mathbf{x}) > 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

b. **negative definite** *if* $Q(\mathbf{x}) < 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

c. **indefinite** *if* $Q(\mathbf{x})$ *takes on both positive and negative values.*

d. **positive semi-definite** *if* $Q(\mathbf{x}) \geq 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

e. **negative semi-definite** *if* $Q(\mathbf{x}) \leq 0$ *for all* $\mathbf{x} \neq \mathbf{0}$.

By extension, a symmetric matrix A is said to be positive definite if the corresponding quadratic form is positive definite. The other terms defined here can be extended to apply to symmetric matrices in a similar way.

Theorem 6.7 *Let A be an $n \times n$ symmetric matrix. The quadratic form $\mathbf{x}^T A\mathbf{x}$ is:*

a. positive definite if and only if all the eigenvalues of A are positive.

b. negative definite if and only if all the eigenvalues of A are negative.

c. indefinite if and only if A has both positive and negative eigenvalues.

Proof There is an orthogonal matrix P such that

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T P^T A P\mathbf{y} = \mathbf{y}^T D\mathbf{y} = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

where $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the eigenvalues of A. Since P is invertible and $\mathbf{x} = P\mathbf{y}$ it follows that $\mathbf{y} \neq \mathbf{0}$ whenever $\mathbf{x} \neq \mathbf{0}$. It should be clear then that the signs of the eigenvalues determine the range of possible values of the quadratic form.

Exercises

1. Eliminate the cross product terms from the following quadratic forms by performing a change of variable:

   (a) $6x_1^2 + 4x_1x_2 + 3x_2^2$

   (b) $4x_1x_2$

   (c) $4x_2^2 + 4x_1x_3$

   (d) $2x_1^2 + 4x_1x_3 + x_2^2 - x_3^2$

2. Find a change of variable that eliminates the cross product terms from the following quadratic form. For what values of a is the quadratic form positive definite? negative definite? indefinite?

$$x_1^2 + x_2^2 + x_3^2 + 2a\,(x_1x_2 + x_1x_3 + x_2x_3)$$

3. Find a change of variable that eliminates the cross product terms from the following quadratic form. For what values of a is the quadratic form positive definite? negative definite? indefinite? (HINT: show that the vectors in the Haar basis for $\mathbb{R}^4$ are eigenvectors of the matrix of this quadratic form.)

$$x_1^2 + x_2^2 + x_3^2 + x_4^2 + 2a\,(x_1x_2 + x_1x_3 + x_1x_4 + x_2x_3 + x_2x_4 + x_3x_4)$$

4. Each of the following equations is of the form $\mathbf{x}^T A\mathbf{x} = c$. Plot each equation and the eigenspaces of the corresponding matrix A. Also find the coordinates of the points on the plot that are closest to and farthest from the origin.

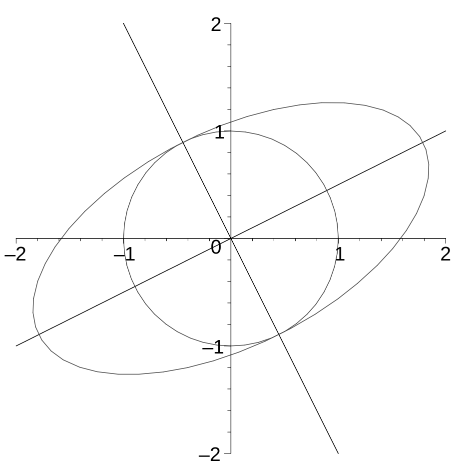
   (a) $6x_1^2 + 4x_1x_2 + 3x_2^2 = 7$ (see 1(a))

   (b) $x_1^2 - 2x_1x_2 + x_2^2 = 4$

   (c) $2x_1x_2 = 1$

   (d) $5x_1^2 + 8x_1x_2 + 5x_2^2 = 1$

5. Let A be a symmetric matrix. You are given the following plot of a unit circle and the circle after it is transformed by multiplying by matrix A.



Draw a picture to illustrate how the unit circle would be transformed if it was multiplied by A^{-1}? by A^2? by A^T?

6. Is the origin a maximum, minimum, or saddle point for the following quadratic forms:
   (a) $Q_1(x) = x_1^2 + 10x_1x_2 + x_2^2$
   (b) $Q_1(x) = x_1^2 - 10x_1x_2 + x_2^2$
7. Let $Q(\mathbf{x}) = x_1^2 + 4x_1x_2 + 3x_2^2$.
   (a) Show that $Q(\mathbf{x})$ is indefinite (despite the fact that all the coefficients are positive).
   (b) Find a specific point where $Q(\mathbf{x})$ is negative.
8. Find the points on $x_1^2 + 4x_1x_2 + 4x_2^2 = 1$ that are closest to and farthest from the origin.
9. For each of the following (a) sketch the surface, and (b) find the points (if any) that are closest to and farthest from the origin.
   (a) $x_1^2 + 4x_2^2 + 9x_3^2 = 1$
   (b) $x_1^2 + 4x_2^2 - 9x_3^2 = 1$
   (c) $x_1^2 - 4x_2^2 - 9x_3^2 = 1$
   (d) $-x_1^2 - 4x_2^2 + 9x_3^2 = 1$
10. (a) For what values of a is the matrix $\begin{bmatrix} a & 3 \\ 3 & a \end{bmatrix}$ positive definite? negative definite? indefinite?
    (b) For what values of a is the matrix $\begin{bmatrix} a & 3 & 3 \\ 3 & a & 3 \\ 3 & 3 & a \end{bmatrix}$ positive definite? negative definite? indefinite?
11. Show that if A is any matrix then the symmetric matrix A^TA is positive semi-definite. Under what conditions will A^TA be positive definite?
12. Let $\mathbf{u}$ be any non-zero vector in $\mathbb{R}^2$, then $\mathbf{u}\mathbf{u}^T$ will be a symmetric 2×2 matrix. Let $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{u}\mathbf{u}^T\mathbf{x}$ be the corresponding quadratic form.
    (a) What are the principal axes of $Q(\mathbf{x})$?
    (b) What is the plot of $Q(\mathbf{x}) = 1$?
    (c) What are the maximum and minimum values of $Q(\mathbf{x})$ subject to the constraint that $\|\mathbf{x}\| = 1$?
13. Let $\mathbf{x}$ be a unit eigenvector of matrix A corresponding to an eigenvalue of 4. What is the value of $\mathbf{x}^TA\mathbf{x}$?
14. Let A be an invertible symmetric matrix. Explain why A^2 must be positive definite.
15. If A and B are positive definite matrices of the same size show that $A + B$ is also positive definite.

Using MAPLE

Example 1.

We will illustrate using **Maple** to plot an equation of the form $Q(\mathbf{x}) = c$ in $\mathbb{R}^3$.
Let $Q(\mathbf{x}) = 5\,x^2 + 8\,y^2 + 5\,z^2 - 2\,xy + 4\,xz - 2\,yz$. The matrix of this quadratic form would be

$$\begin{bmatrix} 5 & -1 & 2 \\ -1 & 8 & -1 \\ 2 & -1 & 5 \end{bmatrix}$$

We begin as follows:

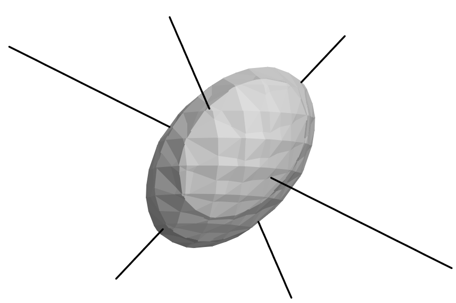
```
>vx:=[x,y,z];
>A:=matrix(3,3,[5,-1,2,-1,8,-1,2,-1,5]):
>ev:=eigenvects(A);

    [3,1,{[1,0,-1]}], [9,1,{[1,-2,1]}], [6,1,{[1,1,1]}]
>qf:=evalm(transpose(vx)&*A&*vx):
```

We now have the basic information we will use for our plots. We will plot the surface correponding to $Q(\mathbf{x}) = 20$ (called p1 below) and the eigenspaces of matrix A (called p1 below).

```
>with(plots):
>p1:=implicitplot3d(q1=20,x=-3..3,y=-3..3,z=-3..3,
                    style=patchnogrid,lightmodel=light2):
>p2:=spacecurve({[t,0,-t],[t,-2*t,t],[t,t,t]},t=-3..3,thickness=3,color=black):
>display([p1,p2]);
```

We then get the following plot:



Try rotating this plot to see the figure from various angles. It should be clear that the eigenspaces are the axes of the ellipsoid.

Example 2.

Let $Q(\mathbf{x}) = 4x_1^2 - 4x_1x_2 - 3x_2^2$. $Q(\mathbf{x})$ is an indefinite quadratic form. If we plotted the surface corresponding to this quadratic form we would get a saddle curve. If we plotted $Q(\mathbf{x}) = 8$ we would get a hyperbola. In this example we will use **Maple** to illustrate the relationship between these two plots.

After we load the plots package we define the quadratic form and call it qf. Next we define a plot of the surface generated by the quadratic form and the plane $z = 8$.

```
>with(plots):
>qf:=4*x1^2-4*x1*x2-3*x2^2;
>p1:=plot3d({qf,8},x=-3..3,y=-3..3):
```

We now want to plot the line of intersection of the plane and the saddle curve.

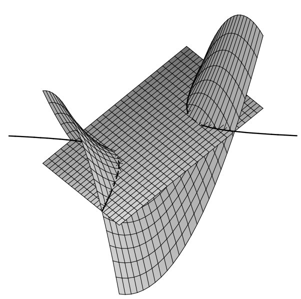
```
>sol:=solve(f=8, x2);
```

$$2/3\,x1 + 2/3\,\sqrt{4\,x1^2 - 6}\,, 2/3\,x1 + 2/3\,\sqrt{4\,x1^2 - 6}$$

This result gives us the the x2 coordinate of the hyperbola in terms of x1. The third coordinate of the line of intersection is the constant 8. So now we plot the two branches that make up this line and display the plots together.

```
>p2:=spacecurve({[x1,sol[1],8],[x1,sol[2],8]},x1=-3..3,color=black,thickness=3):
>display( [p1,p2]);
```

This gives



Example 3.

In $\mathbb{R}^3$ the constraint $\|\mathbf{x}\| = 1$ means that $\mathbf{x}$ lies on the unit sphere. Trying to maximize or minimize a quadratic form under the constraint $\|\mathbf{x}\| = 1$ can therefore be seen as trying to find the maximum or minimum values of the quadratic form on the unit sphere. In this example we will use **Maple** to visualize this problem.

We will use the fact that a point on the unit sphere has the form $[\cos(s)\sin(t), \sin s \sin(t), \cos(t)]$. In the following code: A is a symmetric matrix that defines a quadratic form. qf is the value of the quadratic form on the unit sphere. p1 is a plot of the unit sphere where the color at any point is determined by the value of the quadratic form, where the brightness corresponds to the value of the quadratic form. The brightest point is the maximum, the darkest point is the minimum. p2 is the plot of the eigenspace with the smallest eigenvalue, this axis will pass through the darkest point on the sphere.

```
>x:=[cos(s)*sin(t),sin(s)*sin(t),cos(t)]:
>A:=matrix(3,3,[1.0,2.0,3.0,2.0,0,0,3.0,0,3.0]):
>ev:=eigenvects(A);
     ev := [-2.3669098, 1, {[.70247571, -.59358045, -.39267049]}],
            [.9330141, 1, {[-.36033969, -.77242064, .52299297]}],
            [5.4338956, 1, {[.61374520, .22589510, .75649734]}]
>qf:=evalm(transpose(x)&*A&*x):
>p1:=plot3d(x,s=0..2*Pi,t=0..2*Pi,color=COLOR(RGB,qf,qf,qf),scaling=constrained,style=patchno
>p2:=spacecurve(evalm(t*ev[1][3][1]),t=-1.8..1.8,thickness=2,color=black):
>display([p1,p2],orientation=[-17,75]);
```

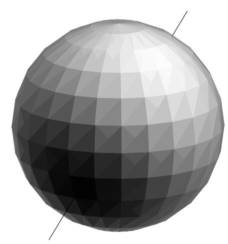


Figure 6.11: Visualizing a quadratic form in $\mathbb{R}^3$.

6.3 Optimization

A type of problem that turns up frequently in many applications is that of finding the maximum or minimum value of a quadratic form $Q(\mathbf{x})$ where $\mathbf{x}$ is constrained to lie in a particular set. It often happens that the problem can be stated in such a way that $\mathbf{x}$ must be a unit vector. This type of problem is called *constrained optimization.*

As we illustrate in the next example, this type of problem is easy to solve if the quadratic form has no cross product terms.

Example 6.3.9

Find the maximum and minimum values of the quadratic form $Q(\mathbf{x}) = 3x_1^2 + 6x_2^2 + 2x_3^2$ subject to the constraint that $\|\mathbf{x}\| = 1$.
We have

$$\begin{aligned} Q(\mathbf{x}) &= 3x_1^2 + 6x_2^2 + 2x_3^2 \\ &\leq 6x_1^2 + 6x_2^2 + 6x_3^2 \\ &= 6(x_1^2 + x_2^2 + x_3^2) \\ &= 6(1) \\ &= 6 \end{aligned}$$

The above shows that the quadratic form always takes on values less than or equal to 6 given the constraint $\|\mathbf{x}\| = 1$. But when $\mathbf{x} = [0, 1, 0]^T$ we have $Q(\mathbf{x}) = 6$ so it follows that 6 is the maximum value of $Q(\mathbf{x})$ given that $\|\mathbf{x}\| = 1$.
For the minimum value we have a similar argument:

$$\begin{aligned} Q(\mathbf{x}) &= 3x_1^2 + 6x_2^2 + 2x_3^2 \\ &\geq 2x_1^2 + 2x_2^2 + 2x_3^2 \\ &= 2(x_1^2 + x_2^2 + x_3^2) \\ &= 2 \end{aligned}$$

So $Q(\mathbf{x})$ is never smaller than 2. Furthermore when $\mathbf{x} = [0, 0, 1]^T$ we have $Q(\mathbf{x}) = 2$ and so 2 is the minimum value of $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = 1$.

Now suppose you want to find the maximum and minimum values of $Q(\mathbf{x}) = 5x_1^2 + 2x_2^2$ subject to the constraint $\|\mathbf{x}\| = 1$. From the previous discussion it should be clear that the maximum value is 5 at the points $(\pm 1, 0)$ and the minimum value is 2 at $(0, \pm 1)$. This quadratic form can be plotted as a surface in 3 dimensions.

If we look at **Figures 6.12** and **6.13** we see that the plot of the quadratic form is a *paraboloid.* The optimization problem is to find the maximum and minimum values on the boundary of the second of the above plots. This boundary is the intersection of the paraboloid and a cylinder of radius 1.

In the next example we will make the sign of the second term negative. In this case it should be clear that for $Q(\mathbf{x}) = 5x_1^2 - 2x_2^2$ the maximum value will again be 5 and the minimum value will be -2. The surface in this case, shown in **Figures 6.14** and **6.15**, is a *hyperbolic paraboloid*, or a *saddle curve.*

Theorem 6.8 *Let $Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x}$ be a quadratic form. The maximum value of $Q(\mathbf{x})$ subject to the constraint that $\|\mathbf{x}\| = 1$ is M, where M is the largest eigenvalue of A. The minimum value of $Q(\mathbf{x})$*

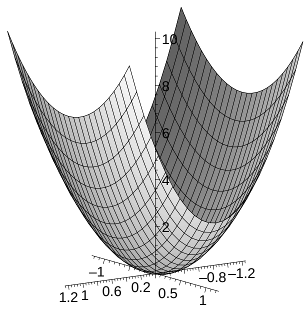


Figure 6.12: $Q(\mathbf{x}) = 5x_1^2 + 2x_2^2$

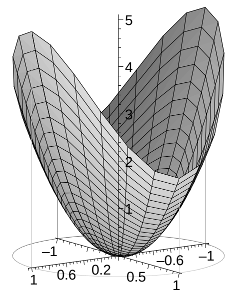


Figure 6.13: $Q(\mathbf{x}) = 5x_1^2 + 2x_2^2$ with boundary $\|\mathbf{x}\| = 1$

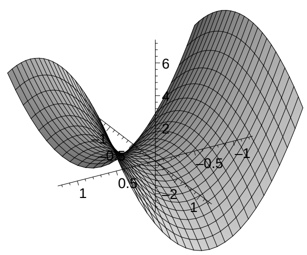


Figure 6.14: $Q(\mathbf{x}) = 5x_1^2 - 2x_2^2$

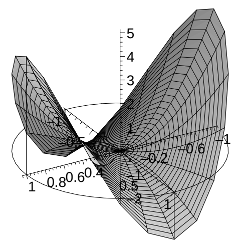


Figure 6.15: $Q(\mathbf{x}) = 5x_1^2 - 2x_2^2$ with boundary $\|\mathbf{x}\| = 1$

subject to the constraint that $\|\mathbf{x}\| = 1$ *is* m*, where* m *is the smallest eigenvalue of* A*. The value of* M *will be taken on when* $\mathbf{x}$ *is a unit eigenvector corresponding to* M*. The value of* m *will be taken on when* $\mathbf{x}$ *is a unit eigenvector corresponding to* m*.*

Proof. We can write

$$\mathbf{x}^T A\mathbf{x} = \mathbf{y}^T D\mathbf{y}$$

where D is diagonal and P is orthogonal and $\mathbf{x} = P\mathbf{y}$. Since P is orthogonal the restriction $\|\mathbf{x}\| = 1$ is equivalent to $\|\mathbf{y}\| = 1$ so it suffices to find the maximum and minimum values of $\mathbf{y}^T D\mathbf{y}$ subject to the constraint $\|\mathbf{y}\| = 1$.

We can assume that D has the form

$$\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ & & \vdots & 0 \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$.

We then have

$$\begin{aligned} \mathbf{y}^T D\mathbf{y} &= \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 \\ &\leq \lambda_1 y_1^2 + \lambda_1 y_2^2 + \cdots + \lambda_1 y_n^2 \\ &= \lambda_1(y_1^2 + y_2^2 + \cdots + y_n^2) \\ &= \lambda_1(1) \\ &= \lambda_1 \end{aligned}$$

So the maximum value of $\mathbf{y}^T D\mathbf{y}$ can't be larger than λ_1. But when $\mathbf{y} = \begin{bmatrix}1 & 0 & 0 & \dots & 0\end{bmatrix}^T$ we get $\mathbf{y}^T D\mathbf{y} = \lambda_1$ so the maximum value of the quadratic form will be $M = \lambda_1$ (the largest eigenvalue of A). This maximum occurs at $\mathbf{y} = \mathbf{e}_1$. But $\mathbf{x} = P\mathbf{y}$ so the value of $\mathbf{x}$ where the maximum is take on is $P\mathbf{e}_1$, and this would be the first column of P which will be a unit eigenvector corresponding to λ_1.

A similar argument can be used for the minimum.

Example 6.3.10

Find the maximum and minimum of $Q(\mathbf{x}) = 3x_1^2 + 2x_1x_2 + 3x_2^2$ subject to the constraint $x_1^2 + x_2^2 = 1$ (i.e., $\|\mathbf{x}\| = 1$).
In this case we have

$$Q(\mathbf{x}) = \mathbf{x}^T A\mathbf{x} = \mathbf{x}^T \begin{bmatrix}3 & 1\\ 1 & 3\end{bmatrix} \mathbf{x}$$

and A has eigenvalues of 4 and 2. So at this point we know the maximum value of the quadratic form will be 4 and the minimum value will be 2. A unit eigenvector for $\lambda = 4$ would be $\mathbf{v}_1 = \begin{bmatrix}1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$ and if we evaluate $Q(\mathbf{x})$ at this point we get

$$\begin{aligned} Q(\mathbf{v}_1) &= 3\left(\frac{1}{\sqrt{2}}\right)^2 + 2\cdot\frac{1}{\sqrt{2}}\frac{1}{\sqrt{2}} + 3\left(\frac{1}{\sqrt{2}}\right)^2 \\ &= \frac{3}{2} + 1 + \frac{3}{2} \\ &= 4 \end{aligned}$$

It is easy to confirm that we also have $Q(-\mathbf{v}_1) = 4$.
Similarly, for $\lambda = 2$ we would have unit eigenvector $\mathbf{v}_2 == \begin{bmatrix}-1/\sqrt{2}\\ 1/\sqrt{2}\end{bmatrix}$ and substituting this into $Q(\mathbf{x})$ we have

$$\begin{aligned} Q(\mathbf{v}_2) &= 3\left(-\frac{1}{\sqrt{2}}\right)^2 + 2\cdot\frac{-1}{\sqrt{2}}\frac{1}{\sqrt{2}} + 3\left(\frac{1}{\sqrt{2}}\right)^2 \\ &= \frac{3}{2} - 1 + \frac{3}{2} \\ &= 2 \end{aligned}$$

In the following plot we see the plot of the ellipses $Q(\mathbf{x}) = 4$, $Q(\mathbf{x}) = 2$, and the unit circle.
The eigenvalues of A are both positive so $Q(\mathbf{x})$ is positive definite. The plot of $Q(\mathbf{x}) = k$ would therefore be an ellipse oriented along the principal axes. These axes are the eigenspaces found above and are independent from k. If $k > 4$ the ellipse would not intersect the unit circle. If $0 < k < 2$ again the ellipse would not intersect the unit circle. The points where the unit circle intersects the ellipse $Q(\mathbf{x}) = 4$ are points on the eigenspace corresponding to $\lambda = 4$. The points where the unit circle intersects the ellipse $Q(\mathbf{x}) = 2$ are points on the eigenspace corresponding to $\lambda = 2$.

Example 6.3.11

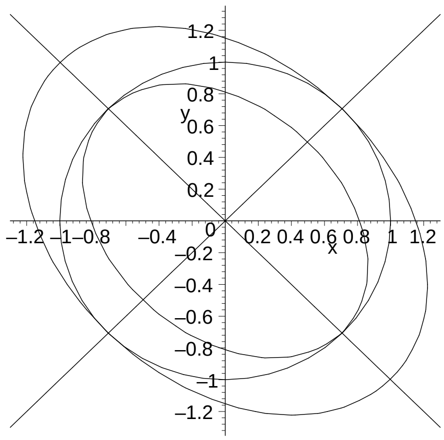


Figure 6.16: Plots of $Q(\mathbf{x}) = 4$, $Q(\mathbf{x}) = 2$, and the unit circle

Let $Q(\mathbf{x}) = 2x_1x_2 + 2x_1x_3 + 2x_2x_3$. Then

$$Q(\mathbf{x}) = \mathbf{x}^T \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \mathbf{x}$$

Matrix A has eigenvalues 2 and -1 (multiplicity 2) and can be orthogonally diagonalized by

$$P = \begin{bmatrix} 1/\sqrt{3} & 1/\sqrt{2} & 1/\sqrt{5} \\ 1/\sqrt{3} & -1/\sqrt{2} & 1/\sqrt{5} \\ 1/\sqrt{3} & 0 & -2/\sqrt{5} \end{bmatrix}$$

If we let $\mathbf{x} = P\mathbf{y}$ would then have $Q(\mathbf{x}) = 2y_1^2 - y_2^2 - y_3^2$.
By the previous theorem, the maximum value of $Q(\mathbf{x})$ would be 2, and there are exactly two unit vectors where this maximum would be taken on, $\mathbf{x} = \pm \begin{bmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{bmatrix}$.

The minimum value would be -1 and this would occur at, for example, $\mathbf{x} = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix}$.

In this case the eigenspace is two dimensional so there are infinitely many unit vectors where the minimum is taken on.

Historical Note

If you have taken a higher level calculus course that dealt with Lagrange multipliers you might be interested in the following. The theory of eigenvalues and eigenvectors grew out of the study of quadratic functions on a sphere. Suppose you have a quadratic

$$f(x, y) = ax^2 + 2bxy + cy^2$$

and you want to maximize or minimize this function on the unit circle, $g(x, y) = x^2 + y^2 - 1 = 0$. Using the method of Lagrange multipliers you would set

$$f_x = \lambda g_x$$

$$f_y = \lambda g_y$$

This would result in the following linear system

$$\begin{aligned} ax + by &= \lambda x \\ bx + cy &= \lambda y \end{aligned}$$

and this can be written

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix} = \lambda \begin{bmatrix} x \\ y \end{bmatrix}$$

The conclusion is then that the maximum and minimum that you are looking for would correspond to the unit eigenvectors of $\begin{bmatrix} a & b \\ b & c \end{bmatrix}$. This is the source of using the symbol λ to stand for eigenvalues.

Exercises

1. Given that there is a change of variable $\mathbf{x} = P\mathbf{y}$ that transforms the quadratic form $Q(\mathbf{x}) = 5x_1^2 + 6x_2^2 + 7x_3^2 + 4x_1x_2 - 4x_2x_3$ into $9y_1^2 + 6y_2^2 + 3y_3^2$, find the maximum and minimum values of $Q(\mathbf{x})$ subject to the constraint that $\|\mathbf{x}\| = 1$ and the find the values of $\mathbf{x}$ where they are taken on.

2. Given that there is a change of variable $\mathbf{x} = P\mathbf{y}$ that transforms the quadratic form $Q(\mathbf{x}) = 3x_1^2 + 2x_2^2 + 2x_3^2 + 2x_1x_2 + 2x_1x_3 + 4x_2x_3$ into $5y_1^2 + 2y_2^2$, find the maximum and minimum values of $Q(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$ and find the values of $\mathbf{x}$ where they are taken on.

3. Find the maximum value of the following quadratic forms, $Q(\mathbf{x})$, subject to the constraint that $\|\mathbf{x}\| = 1$.
   (a) $Q(\mathbf{x}) = 3x_1^2 - 2x_1x_2 + 7x_2^2$.
   (b) $Q(\mathbf{x}) = 5x_1^2 - 2x_1x_2 - 3x_2^2$.

4. Find the maximum value of $Q(\mathbf{x})$ subject to the constraint that $\|\mathbf{x}\| = 1$.
   (a) $Q(\mathbf{x}) = x_1^2 + 4x_2^2 + 2x_3^2$
   (b) $Q(\mathbf{x}) = 5x_1^2 - 4x_1x_2 + 5x_2^2$
   (c) $Q(\mathbf{x}) = 7x_1^2 + 3x_1x_2 + 3x_2^2$

5. Let $Q(\mathbf{x}) = 18x_1^2 + 58x_3^2 + 54x_1x_2 + 18x_1x_3 + 114x_2x_3$.
   (a) Write $Q(\mathbf{x})$ in the form $\mathbf{x}^TA\mathbf{x}$.
   (b) Given that matrix A has eigenvalues 99, 18 and -41 what is the maximum value of $Q(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$.
   (c) Find a unit vector $\mathbf{v}_1$ where $Q(\mathbf{x})$ takes on its maximum value.
   (d) Find the maximum value of $Q(\mathbf{x})$ subject to the constraints $\|\mathbf{x}\| = 1$ and $\mathbf{x}^T\mathbf{v}_1 = 0$. (Hint: try to interpret these constraints geometrically.)

6. Let $Q(\mathbf{x}) = 2x_1x_2 - 6x_1x_3 + 2x_2x_3$.
   (a) Find the maximum value of $Q(\mathbf{x})$ subject to $\|\mathbf{x}\| = 1$.
   (b) Find a unit vector $\mathbf{v}_1$ where the maximum value is taken on.
   (c) Find the maximum value of $Q(\mathbf{x})$ subject to the constraints $\|\mathbf{x}\| = 1$ and $\mathbf{x}^T\mathbf{v}_1 = 0$. (Hint: try to interpret these constraints geometrically.)

7. Suppose A is a 3×3 symmetric matrix with eigenvalues .2, 1.5, and 2 find:
   (a) the maximum value of $\mathbf{x}^TA\mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.
   (b) the maximum value of $\mathbf{x}^TA^2\mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.
   (c) the maximum value of $\mathbf{x}^TA^{-1}\mathbf{x}$ subject to $\|\mathbf{x}\| = 1$.

8. Suppose $Q(\mathbf{x})$ is a quadratic form with maximum value M when $\mathbf{x}$ is subject to the constraint $\|\mathbf{x}\| = 1$. What is the maximum value of $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = 2$? What is the maximum value of $Q(\mathbf{x})$ subject to the constraint $\|\mathbf{x}\| = k$