Chapter 3

Linear Algebra

As the reader is aware by now, calculus has its roots in physics and has become a very useful tool for modelling the physical world. Another very important area of mathematics is linear algebra. Physics students who have taken a course in linear algebra in a mathematics department might not come away with this perception. It is not until students take more advanced classes in physics that they begin to realize that a good grounding in linear algebra can lead to a better understanding of the behavior of physical systems.

In this chapter we will introduce some of the basics of linear algebra for finite dimensional vector spaces and we will reinforce these concepts through generalizations in later chapters to infinite dimensional vector spaces. In keeping with the theme of our text, we will apply some of these ideas to the coupled systems introduced in the last chapter.

3.1 Vector Spaces

Much of the discussion and terminology that we will use comes from the theory of vector spaces. Until now you may only have dealt with finite dimensional vector spaces. Even then, you might only be comfortable with vectors in two and three dimensions. We will review a little of what we know about finite dimensional vector spaces. In later sections we will introduce the more general function spaces, which will be useful in later application in the text.

The notion of a vector space is a generalization of the three dimensional vector spaces that you have seen in introductory physics and calculus. In three dimensions, we have objects called vectors, which you first visualized as arrows of a specific length and pointing in a given direction. To each vector, we can associate a point in a three dimensional Cartesian system. We just attach the tail of the vector $\mathbf{v}$ to the origin and the head lands at some point, (x, y, z). We then used the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ along the coordinate axes to write the vector in the form

$$\mathbf{v} = x\mathbf{i} + y\mathbf{j} + z\mathbf{k}.$$

Having defined vectors, we then learned how to add vectors and multiply vectors by numbers, or scalars. Under these operations, we expected to get back new vectors. Then we learned that there were two types of multiplication of vectors. We could multiply two vectors to get either a scalar or a vector. This lead to the operations of dot and cross products, respectively. The dot product was useful for determining the length of a vector, the angle between two vectors, or if the vectors were orthogonal.

In physics you first learned about vector products when you defined work, $W = \underline{\mathrm{F}} \cdot \mathbf{r}$. Cross products were useful in describing things like torque, $\tau = \mathbf{r} \times \mathbf{F}$, or the force on a moving charge in a magnetic field, $\mathbf{F} = q\mathbf{v} \times \mathbf{B}$. We will return to these more complicated vector operations later when reviewing Maxwell's equations of electrodynamics.

The basic concept of a vector can be generalized to spaces of more than three dimensions. You may first have seen this in your linear algebra class. The properties outlined roughly above need to be preserved. So, we have to start with a space of vectors and the operations between them. We also need a set of scalars, which generally come from some *field*. However, in our applications the field will either be the set of real numbers or the set of complex numbers.

A *vector space* V over a field F is a set that is closed under addition and scalar multiplication and satisfies the following conditions: For any $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$ and $a, b \in F$

1. $\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.

2. $(\mathbf{u}+\mathbf{v})+\mathbf{w}=\mathbf{u}+(\mathbf{v}+\mathbf{w})$.
3. There exists a $\mathbf{0}$ such that $\mathbf{0}+\mathbf{v}=\mathbf{v}$.
4. There exists a $-\mathbf{v}$ such that $\mathbf{v}+(-\mathbf{v})=0$.
5. $a(b\mathbf{v})=(ab)\mathbf{v}$.
6. $(a+b)\mathbf{v}=a\mathbf{v}+b\mathbf{v}$.
7. $a(\mathbf{u}+\mathbf{v})=a\mathbf{u}+b\mathbf{v}$.
8. $1(\mathbf{v})=\mathbf{v}$.

In three dimensions the unit vectors $\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$ play an important role. Any vector in the three dimensional space can be written as a linear combination of these vectors,

$$\mathbf{v}=x\mathbf{i}+y\mathbf{j}+z\mathbf{k}.$$

In fact, given any three non-coplanar vectors, $\{\mathbf{a}_1,\mathbf{a}_2,\mathbf{a}_3\}$, all vectors can be written as a linear combination of those vectors,

$$\mathbf{v}=c_1\mathbf{a}_1+c_2\mathbf{a}_1+c_3\mathbf{a}_1.$$

Such vectors are said to span the space and are called a basis for the space.

We can generalize these ideas. In an n-dimensional vector space any vector in the space can be represented as the sum over n linearly independent vectors (the equivalent of non-coplanar vectors). Such a *linearly independent* set of vectors $\{\mathbf{v}_j\}_{j=1}^n$ satisfies the condition

$$\sum_{j=1}^{n} c_j\mathbf{v}_j=\mathbf{0} \quad \Leftrightarrow \quad c_j=0.$$

Note that we will often use summation notation instead of writing out all of the terms in the sum.

The *standard basis* in an n-dimensional vector space is a generalization of the standard basis in three dimensions ($\mathbf{i}$, $\mathbf{j}$ and $\mathbf{k}$). We define

$$\mathbf{e}_k=(0,\ldots,0,\underbrace{1}_{k\text{th space}},0,\ldots,0),\quad k=1,\ldots,n. \tag{3.1}$$

Then, we can expand any $\mathbf{v} \in V$ as

$$\mathbf{v} = \sum_{k=1}^{n} v_k \mathbf{e}_k, \tag{3.2}$$

where the v_k's are called the components of the vector in this basis. Sometimes we will write $\mathbf{v}$ as an n-tuple $(v_1, v_2, \ldots, v_n)$. This is similar to the ambiguous use of (x, y, z) to denote both vectors in as well as to represent points in the three dimensional space.

The only other thing we will need at this point is to generalize the dot product. Recall that there are two forms for the dot product in three dimensions. First, one has that

$$\mathbf{u} \cdot \mathbf{v} = uv \cos\theta, \tag{3.3}$$

where u and v denote the length of the vectors. The other form is the component form:

$$\mathbf{u} \cdot \mathbf{v} = u_1 v_1 + u_2 v_2 + u_3 v_3 = \sum_{k=1}^{3} u_k v_k. \tag{3.4}$$

Of course, this form is easier to generalize. So, we define the *scalar product* between two n-dimensional vectors as

$$< \mathbf{u}, \mathbf{v} >= \sum_{k=1}^{n} u_k v_k. \tag{3.5}$$

Actually, there are a number of notations that are used in other texts. One can write the scalar product as $(\mathbf{u}, \mathbf{v})$ or even in the Dirac bra-ket notation $< \mathbf{u}|\mathbf{v} >$.

We note that the (real) scalar product satisfies some simple properties. For vectors $\mathbf{v}$, $\mathbf{w}$ and real scalar α we have

1. $< \mathbf{v}, \mathbf{v} >\geq 0$ and $< \mathbf{v}, \mathbf{v} >= 0$ if and only if $\mathbf{v} = \mathbf{0}$.

2. $< \mathbf{v}, \mathbf{w} >=< \mathbf{w}, \mathbf{v} >$.

3. $< \alpha\mathbf{v}, \mathbf{w} >= \alpha < \mathbf{v}, \mathbf{w} >$.

While it does not always make sense to talk about angles between general vectors in higher dimensional vector spaces, there is one concept that is useful. It is that of orthogonality, which in three dimensions is another way of saying the vectors are perpendicular to each other. So, we also say that vectors $\mathbf{u}$ and $\mathbf{v}$ are *orthogonal* if and only if $<\mathbf{u},\mathbf{v}>=0$. If $\{\mathbf{a}_k\}_{k=1}^n$, is a set of basis vectors such that

$$<\mathbf{a}_j,\mathbf{a}_k>=0, \quad k\neq j,$$

then it is called an *orthogonal basis.*

If in addition each basis vector is a unit vector, then one has an *orthonormal basis.* This generalization of the unit basis can be expressed more compactly. We will denote such a basis of unit vectors by $\mathbf{e}_j$ for $j=1\dots n$. Then,

$$<\mathbf{e}_j,\mathbf{e}_k>=\delta_{jk}, \tag{3.6}$$

where we have introduced the Kronecker delta

$$\delta_{jk}\equiv\begin{cases}0, & j\neq k\\ 1, & j=k\end{cases} \tag{3.7}$$

The process of making basis vectors have unit length is called *normalization.* This is simply done by dividing by the *length* of the vector. Recall that the length of a vector, $\mathbf{v}$, is obtained as $v=\sqrt{\mathbf{v}\cdot\mathbf{v}}$. So, if we want to find a unit vector in the direction of $\mathbf{v}$, then we simply normalize it as

$$\hat{\mathbf{v}}=\frac{\mathbf{v}}{v}.$$

Notice that we used a hat to indicate that we have a unit vector. Furthermore, if $\{\mathbf{a}_j\}_{j=1}^n$, is a set of orthogonal basis vectors, then

$$\mathbf{e}_j=\frac{\mathbf{a}_i}{\sqrt{<\mathbf{a}_j,\mathbf{a}_j>}}, \quad j=1\dots n.$$

Let $\{\mathbf{a}_k\}_{k=1}^n$, be a set of orthogonal basis vectors for vector space V. We know that any vector $\mathbf{v}$ can be represented in terms of this basis, $\mathbf{v}=\sum_{k=1}^n v_k\mathbf{a}_k$. If we know the basis and vector, can we find the components? The answer is yes. We can use the scalar product of $\mathbf{v}$ with each basis element $\mathbf{a}_j$. Using the properties of the scalar product, we have for $j = 1, \ldots, n$

$$\begin{aligned} < \mathbf{a}_j, \mathbf{v} > &= < \mathbf{a}_j, \sum_{k=1}^{n} v_k \mathbf{a}_k > \\ &= \sum_{k=1}^{n} v_k < \mathbf{a}_j, \mathbf{a}_k > . \end{aligned} \tag{3.8}$$

Since we know the basis elements, we can easily compute the numbers

$$A_{jk} \equiv < \mathbf{a}_j, \mathbf{a}_k >$$

and

$$b_j \equiv < \mathbf{a}_j, \mathbf{v} > .$$

Therefore, the system (3.8) for the v_k's is a linear algebraic system, which takes the form

$$b_j = \sum_{k=1}^{n} A_{jk} v_k. \tag{3.9}$$

We can write this set of equations in a more compact form. The set of numbers A_{jk}, $j, k = 1, \ldots n$ are the elements of an $n \times n$ matrix A with A_{jk} being an element in the jth row and kth column. Also, v_j and b_j can be written as column vectors, $\mathbf{v}$ and $\mathbf{b}$, respectively. Thus, system (3.8) can be written in matrix form as

$$A\mathbf{v} = \mathbf{b}.$$

However, if the basis is orthogonal, then the matrix $A_{jk} \equiv < \mathbf{a}_j, \mathbf{a}_k >$ is diagonal and the system is easily solvable. Recall that two vectors are orthogonal if and only if

$$< \mathbf{a}_i, \mathbf{a}_j > = 0, \quad i \neq j. \tag{3.10}$$

Thus, in this case we have that

$$< \mathbf{a}_j, \mathbf{v} > = v_j < \mathbf{a}_j, \mathbf{a}_j >, \quad j = 1, \ldots, n. \tag{3.11}$$

or

$$v_j = \frac{< \mathbf{a}_j, \mathbf{v} >}{< \mathbf{a}_j, \mathbf{a}_j >}. \tag{3.12}$$

In fact, if the basis is orthonormal, i.e., the basis consists of an orthogonal set of unit vectors, then A is the identity matrix and the solution takes on a simpler form:

$$v_j = < \mathbf{a}_j, \mathbf{v} > . \tag{3.13}$$

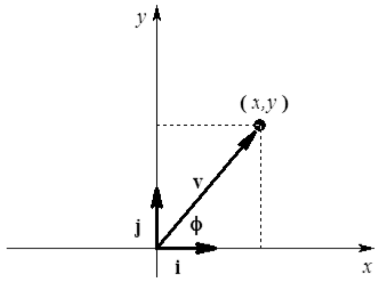

Figure 3.1: Vector **v** in a standard coordinate system.

3.2 Linear Transformations

A main theme in linear algebra is to study linear transformations between vector spaces. These come in many forms and there are an abundance of applications in physics. For example, the transformation between the spacetime coordinates of observers moving in inertial frames in the theory of special relativity constitute such a transformation.

A simple example often encountered in physics courses is the rotation by a fixed angle. This is the description of points in space using two different coordinate bases, one just a rotation of the other by some angle. We begin with a vector **v** as described by a set of axes in the standard orientation, as shown in Figure 3.1. Also displayed in this figure are the unit vectors. To find the coordinates (x, y), one needs only draw a perpendicular to the axes and read the coordinate off the axis.

In order to derive the needed transformation we will make use of polar coordinates. In Figure 3.1 we see that the vector makes an angle of ϕ with respect to the positive x-axis. The components (x, y) of the vector can be determined from this angle and the magnitude of **v** as

$$\begin{aligned} x &= v\cos\phi \\ y &= v\sin\phi. \end{aligned} \tag{3.14}$$

We now consider another set of axes at an angle of θ to the old. Such a system is shown in Figure 3.2. We will designate these axes as x' and y'. Note that the basis vectors are different in this system. Projections to the axes are shown. Comparing the coordinates in both systems shown in

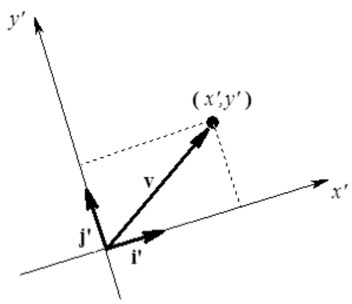

Figure 3.2: Vector **v** in a rotated coordinate system.

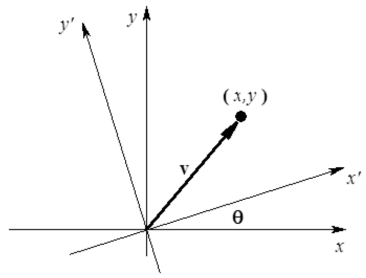

Figure 3.3: Comparison of the coordinate systems.

Figures 3.1-3.2, we see that the primed coordinates are not the same as the unprimed ones.

In Figure 3.3 the two systems are superimposed on each other. The polar form for the primed system is given by

$$\begin{aligned} x' &= v\cos(\phi+\theta) \\ y' &= v\sin(\phi+\theta). \end{aligned} \tag{3.15}$$

We can use this form to find a relationship between the two systems. Namely, we use the addition formula for trigonometric functions to obtain

$$\begin{aligned} x' &= v\cos\phi\cos\theta - v\sin\phi\sin\theta \\ y' &= v\sin\phi\cos\theta + v\cos\phi\sin\theta. \end{aligned} \tag{3.16}$$

Noting that these expressions involve products of v with $\cos\phi$ and $\sin\phi$, we can use the polar form for x and y to find the desired form:

$$x' = x\cos\theta - y\sin\theta$$

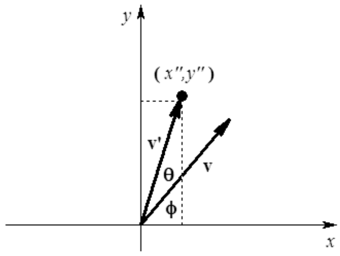

Figure 3.4: Rotation of vector $\mathbf{v}$

$$y' = x\sin\theta + y\cos\theta. \tag{3.17}$$

This is an example of a transformation between two coordinate systems. It is called a rotation by θ. We can designate it generally by

$$(x', y') = \hat{R}_\theta(x, y).$$

It is referred to as a passive transformation, because it does not affect the vector.

An active rotation is one in which one rotates the vector, such as shown in Figure 3.4. One can derive a similar transformation for how the coordinate of the vector change under such a transformation. Denoting the new vector as $\mathbf{v}'$ with new coordinates (x'', y''), we have

$$\begin{aligned} x'' &= x\cos\theta + y\sin\theta \\ y'' &= -x\sin\theta + y\cos\theta. \end{aligned} \tag{3.18}$$

We note that the active and passive rotations are related. Namely,

$$(x'', y'') = \hat{R}_{-\theta}(x, y) = R_\theta(x, y).$$

3.3 Matrices

Linear transformations such as the rotation in the last section can be represented by matrices. Such matrix representations often become the core of a linear algebra class to the extent that one loses sight of their meaning. We will review matrix representations and show how they are useful in solving coupled systems of differential equations later in the chapter.

We begin with the rotation transformation as applied to a vector in Equation (3.18). We write vectors like $\mathbf{v}$ as a column matrix

$$\mathbf{v} = \begin{pmatrix} x \\ y \end{pmatrix}.$$

We can also write the trigonometric functions in a 2×2 matrix form as

$$R_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

Then, the transformation takes the form

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix}. \tag{3.19}$$

This can be written in the more compact form

$$\mathbf{v}' = R_\theta \mathbf{v}.$$

In using the matrix form of the transformation, we have employed the definition of matrix multiplication. Namely, we have multiplied a 2×2 matrix times a 2×1 matrix. (Note that an $n \times m$ matrix has n rows and m columns.) The multiplication proceeds by selecting the ith row of the first matrix and the jth column of the second matrix. Multiply corresponding elements of each and add them. Then, place the result into the ijth entry of the product matrix. This operation can only be performed if the number of columns of the first matrix is the same as the number of columns of the second matrix.

As an example, we multiply a 3×2 matrix times a 2×2 matrix to obtain a 3×2 matrix:

$$\begin{aligned} \begin{pmatrix} 1 & 2 \\ 5 & -1 \\ 3 & 2 \end{pmatrix} \begin{pmatrix} 3 & 2 \\ 1 & 4 \end{pmatrix} &= \begin{pmatrix} 1(3)+2(1) & 1(2)+2(4) \\ 5(3)+(-1)(1) & 5(2)+(-1)(4) \\ 3(3)+2(1) & 3(2)+2(4) \end{pmatrix} \\ &= \begin{pmatrix} 5 & 10 \\ 14 & 6 \\ 11 & 14 \end{pmatrix} \end{aligned} \tag{3.20}$$

In the above example, we have the row $\cos\theta, \sin\theta$ and column x, y. Combining these we obtain $x\cos\theta + y\sin\theta$. This is x'. We perform the same operation for the second row:

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x\cos\theta + y\sin\theta \\ -x\sin\theta + y\cos\theta \end{pmatrix}. \qquad (3.21)$$

In the last section we also introduced active rotations. These were rotations of vectors keeping the coordinate system fixed. Thus, we start with a vector $\mathbf{v}$ and rotate it by θ to get a new vector $\mathbf{u}$. That transformation can be written as

$$\mathbf{u} = \hat{R}_\theta \mathbf{v}, \qquad (3.22)$$

where

$$\hat{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}.$$

Now consider a rotation by $-\theta$. Due to the symmetry properties of the sines and cosines, we have

$$\hat{R}_{-\theta} = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix}.$$

We see that if the 12 and 21 elements of this matrix are interchanged we recover $\hat{R}_\theta$. This is an example of what is called the *transpose* of $\hat{R}_\theta$. Given a matrix, A, its transpose A^T is the matrix obtained by interchanging the rows and columns of A. Formally, let A_{ij} be the elements of A. Then

$$A^T_{ij} = A_{ji}.$$

It is also the case that these matrices are inverses of each other. We can understand this in terms of the nature of rotations. We first rotate the vector by θ as $\mathbf{u} = \hat{R}_\theta \mathbf{v}$ and then rotate $\mathbf{u}$ by $-\theta$ obtaining $\mathbf{w} = \hat{R}_{-\theta}\mathbf{u}$. Thus, the "composition" of these two transformations leads to

$$\mathbf{w} = \hat{R}_{-\theta}\mathbf{u} = \hat{R}_{-\theta}(\hat{R}_\theta \mathbf{v}). \qquad (3.23)$$

We can view this as a net transformation from $\mathbf{v}$ to $\mathbf{w}$ given by

$$\mathbf{w} = (\hat{R}_{-\theta}\hat{R}_\theta)\mathbf{v},$$

where the transformation matrix for the composition is given by $\hat{R}_{-\theta}\hat{R}_\theta$. Actually, if you think about it, we should end up with the original vector. We can compute the resulting matrix by carrying out the multiplication. We obtain

$$\hat{R}_{-\theta}\hat{R}_\theta = \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{3.24}$$

This is the 2×2 identity matrix. We note that the product of these two matrices yields the identity. This is like the multiplication of numbers. If $ab = 1$, then a and b are multiplicative inverses of each other. So, we see here that $\hat{R}_\theta$ and $\hat{R}_{-\theta}$ are inverses of each other as well. In fact, we have determined that

$$\hat{R}_{-\theta} = \hat{R}_\theta^{-1} = \hat{R}_\theta^T, \tag{3.25}$$

where the T designates the transpose. We note that matrices satisfying the relation $A^T = A^{-1}$ are called *orthogonal matrices.*

We can generalize what we have seen with this simple example. We begin with a vector $\mathbf{v}$ in an n-dimensional vector space. We can consider a transformation L that takes $\mathbf{v}$ into a new vector $\mathbf{u}$ as

$$\mathbf{u} = L(\mathbf{v}).$$

We will restrict ourselves to linear transformations. A *linear transformation* satisfies the following condition:

$$L(\alpha\mathbf{a} + \beta\mathbf{b}) = \alpha L(\mathbf{a}) + \beta L(\mathbf{b}) \tag{3.26}$$

for any vectors $\mathbf{a}$ and $\mathbf{b}$ and scalars α and β.

Such linear transformations can be represented by matrices. Take any vector $\mathbf{v}$. It can be represented in terms of a basis. Let's use the standard basis $\{\mathbf{e}_i\}$, $i = 1, \ldots n$. Then we have

$$\mathbf{v} = \sum_{i=1}^n v_i\mathbf{e}_i.$$

Now consider the effect of the transformation L on $\mathbf{v}$, using the linearity property:

$$L(\mathbf{v}) = L\left(\sum_{i=1}^n v_i\mathbf{e}_i\right) = \sum_{i=1}^n v_i L(\mathbf{e}_i). \tag{3.27}$$

Thus, we see that determining how L acts on $\mathbf{v}$ requires that we know how L acts on the basis vectors. Namely, we need $L(\mathbf{e}_i)$. Since $\mathbf{e}_i$ is a vector, this produces another vector in the space. But the resulting vector can be expanded in the basis. Let's assume that the resulting vector takes the form

$$L(\mathbf{e}_i) = \sum_{j=1}^{n} L_{ji}\mathbf{e}_j, \tag{3.28}$$

where L_{ji} is the jth component of $L(\mathbf{e}_i)$ for each $i = 1, \ldots, n$. The matrix of L_{ji}'s is called the matrix representation of the operator L.

Typically, in a linear algebra class you start with matrices and do not see this connection to linear operators. However, there will be times that you will need this connection to understand why matrices are involved. Furthermore, the matrix representation depends on the basis used. We used the standard basis above. However, you could have started with a different basis, such as dictated by another coordinate system. We will not go further into this point at this time and just stick with the standard basis.

Now that we know how L acts on basis vectors, what does this have to say about how L acts on any other vector in the space? We insert expression (3.28) into Equation (3.27). Then we find

$$\begin{aligned} L(\mathbf{v}) &= \sum_{i=1}^{n} v_i L(\mathbf{e}_i) \\ &= \sum_{i=1}^{n} v_i \left(\sum_{j=1}^{n} L_{ji}\mathbf{e}_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{i=1}^{n} v_i L_{ji} \right) \mathbf{e}_j. \end{aligned} \tag{3.29}$$

Since $L(\mathbf{v}) = \mathbf{u}$, we see that the jth component of $\mathbf{u}$ can be written as

$$u_j = \sum_{i=1}^{n} L_{ji} v_i, \quad j = 1 \ldots n. \tag{3.30}$$

This equation can be written in matrix form as

$$\mathbf{u} = L\mathbf{v},$$

where L now takes the role of a matrix. It is similar to the multiplication of the rotation matrix times a vector as seen in the last section. We will just work with matrix representations from here on.

Next, we can compose transformations like we had done with the two rotation matrices. Let $\mathbf{u} = A(\mathbf{v})$ and $\mathbf{w} = B(\mathbf{u})$ for two transformations A and B. (Thus, $\mathbf{v} \to \mathbf{u} \to \mathbf{w}$.) Then a composition of these transformations is given by

$$\mathbf{w} = B(\mathbf{u}) = B(A\mathbf{v}).$$

This can be viewed as a transformation from $\mathbf{v}$ to $\mathbf{w}$ as

$$\mathbf{w} = BA(\mathbf{v}),$$

where the matrix representation of BA is given by the product of the matrix representations of A and B.

To see this, we look at the ijth element of the matrix representation of BA. We first note that the transformation from $\mathbf{v}$ to $\mathbf{w}$ is given by

$$w_i = \sum_{j=1}^{n} (BA)_{ij} v_j. \tag{3.31}$$

However, if we use the successive transformations, we have

$$\begin{aligned} w_i &= \sum_{k=1}^{n} B_{ik} u_k \\ &= \sum_{k=1}^{n} B_{ik} \left(\sum_{j=1}^{n} A_{kj} v_j \right) \\ &= \sum_{j=1}^{n} \left(\sum_{k=1}^{n} B_{ik} A_{kj} \right) v_j. \end{aligned} \tag{3.32}$$

We have two expressions for w_i as sums over v_j. So, the coefficients must be equal. This leads to our result:

$$(BA)_{ij} = \sum_{k=1}^{n} B_{ik} A_{kj}. \tag{3.33}$$

Thus, we have found the component form of matrix multiplication, which resulted from the composition of two linear transformations. This agrees with our earlier example of matrix multiplication: The ij-th component of the product is obtained by multiplying elements in the ith row of B and the jth column of A and summing.

There are many other properties of matrices and types of matrices that one will encounter. We will list a few.

First of all, there is the $n \times n$ *identity matrix*, I. The identity is defined as that matrix satisfying

$$IA = AI = A \tag{3.34}$$

for any $n \times n$ matrix A. The $n \times n$ identity matrix takes the form

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \vdots & 1 \end{pmatrix} \tag{3.35}$$

A component form is given by the Kronecker delta. Namely, we have that

$$I_{ij} = \delta_{ij} \equiv \begin{cases} 0, & i \neq j \\ 1, & i = j \end{cases} \tag{3.36}$$

The *inverse* of matrix A is that matrix A^{-1} such that

$$AA^{-1} = A^{-1}A = I. \tag{3.37}$$

While there is a systematic method for determining the inverse in terms of cofactors, we will not cover it here. It suffices to note that the inverse of a 2×2 matrix is easily obtained. Let

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

Now consider the matrix

$$B = \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

Multiplying these matrices, we find that

$$AB = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix} = \begin{pmatrix} ad-bc & 0 \\ 0 & ad-bc \end{pmatrix}.$$

This is not quite the identity, but it is a multiple of the identity. We just need to divide by $ad - bc$. So, we have found the inverse matrix:

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

We leave it to the reader to show that $A^{-1}A = I$.

The factor $ad - bc$ is the difference in the products of the diagonal and off-diagonal elements of matrix A. This factor is called the *determinant* of A. It is denoted as $\det(A)$ or $|A|$. Thus, we define

$$\det(A) = \begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc. \tag{3.38}$$

For higher dimensional matrices one can write the definition of the determinant. We will for now just indicate the process for 3×3 matrices. We write matrix A as

$$A = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}. \tag{3.39}$$

The determinant of A can be computed in terms of simpler 2×2 determinants. We define

$$\begin{aligned} \det A &= \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} \\ &= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix}. \end{aligned} \tag{3.40}$$

There are many other properties of determinants. For example, if $\det A = 0$, A is called a *singular* matrix. Otherwise, it is called *nonsingular*. If two rows, or columns, of a matrix are multiples of each other, then $\det A = 0$.

A standard application of determinants is the solution of a system of linear algebraic equations using Cramer's Rule. As an example, we consider a simple system of two equations and two unknowns. Let's consider this system of two equations and two unknowns, x and y, in the form

$$\begin{aligned} ax + by &= e, \\ cx + dy &= f. \end{aligned} \tag{3.41}$$

The standard way to solve this is to eliminate one of the variables. (Just imagine dealing with a bigger system!). So, we can eliminate the x's. Multiply the first equation by c and the second equation by a and subtract. We then get

$$(bc - ad)y = (ec - fa).$$

If $bc - ad \neq 0$, then we can solve to y, getting

$$y = \frac{ec - fa}{bc - ad}$$

. Similarly, we find

$$x = \frac{ed - bf}{ad - bc}.$$

We note the the denominators can be replaced with the determinant of the matrix of coefficients,

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

In fact, we can also replace each numerator with a determinant. Thus, our solutions may be written as

$$\begin{aligned} x &= \frac{\begin{vmatrix} e & b \\ f & d \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}} \\ y &= \frac{\begin{vmatrix} a & e \\ c & f \end{vmatrix}}{\begin{vmatrix} a & b \\ c & d \end{vmatrix}}. \end{aligned} \tag{3.42}$$

This is Cramer's Rule for writing out solutions of systems of equations. Note that each variable is determined by placing a determinant with e and

f placed in the column of the coefficient matrix corresponding to the order of the variable in the equation. The denominator is the determinant of the coefficient matrix. This construction is easily extended to larger systems of equations.

Another operation that we have seen earlier is the transpose of a matrix. The transpose of a matrix is a new matrix in which the rows and columns are interchanged. If write an $n \times m$ matrix A in standard form as

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \dots & a_{nm} \end{pmatrix}, \tag{3.43}$$

then the transpose is defined as

$$A^T = \begin{pmatrix} a_{11} & a_{21} & \dots & a_{1n} \\ a_{12} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1m} & a_{2m} & \dots & a_{mn} \end{pmatrix}. \tag{3.44}$$

In index form, we have

$$(A^T)_{ij} = A_{ji}, \quad i, j = 1, \dots, n.$$

As we had seen in the last section, a matrix satisfying

$$A^T = A^{-1}, \quad \text{or} \quad AA^T = A^T A = I,$$

is called an orthogonal matrix. One also can show that

$$(AB)^T = B^T A^T.$$

Finally, the *trace* of a square matrix is the sum of its diagonal elements:

$$\text{Tr}(A) = a_{11} + a_{22} + \dots + a_{nn} = \sum_{i=1}^{n} a_{ii}.$$

We can show that for two square matrices

$$\text{Tr}(AB) = \text{Tr}(BA).$$

3.4 Eigenvalue Problems

3.4.1 An Introduction to Coupled Systems

Recall that one of the reasons we have seemingly digressed into topics in linear algebra and matrices is to solve a coupled system of differential equations. The simplest example is a system of linear differential equations of the form

$$\begin{aligned} \frac{dx}{dt} &= ax + by \\ \frac{dy}{dt} &= cx + dy. \end{aligned} \tag{3.45}$$

We note that this system is coupled. We cannot solve either equation without knowing either $x(t)$ or $y(t)$. A much easier problem would be to solve an uncoupled system like

$$\begin{aligned} \frac{dx}{dt} &= \lambda_1 x \\ \frac{dy}{dt} &= \lambda_2 y. \end{aligned} \tag{3.46}$$

The solutions are quickly found to be

$$\begin{aligned} x(t) &= c_1 e^{\lambda_1 t}, \\ y(t) &= c_2 e^{\lambda_2 t}. \end{aligned} \tag{3.47}$$

Here c_1 and c_2 are two arbitrary constants.

We can determine particular solutions of the system by specifying $x(t_0) = x_0$ and $y(t_0) = y_0$ at some time t_0. Thus,

$$\begin{aligned} x(t) &= x_0 e^{\lambda_1 t}, \\ y(t) &= y_0 e^{\lambda_2 t}. \end{aligned} \tag{3.48}$$

Wouldn't it be nice if we could transform the more general system into one that is not coupled? Let's write our systems in more general form. We write the coupled system as

$$\frac{d}{dt}\mathbf{x} = A\mathbf{x}$$

and the uncoupled system as

$$\frac{d}{dt}\mathbf{y} = \Lambda\mathbf{y},$$

where

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{pmatrix}.$$

We note that Λ is a diagonal matrix.

Now, we seek a transformation between $\mathbf{x}$ and $\mathbf{y}$ that will transform the coupled system into the uncoupled system. Thus, we define the transformation

$$\mathbf{x} = S\mathbf{y}. \tag{3.49}$$

Inserting this transformation into the coupled system we have

$$\begin{aligned} \frac{d}{dt}\mathbf{x} &= A\mathbf{x} \quad \Rightarrow \\ \frac{d}{dt}S\mathbf{y} &= AS\mathbf{y} \quad \Rightarrow \\ S\frac{d}{dt}\mathbf{y} &= AS\mathbf{y}. \end{aligned} \tag{3.50}$$

Multiply both sides by S^{-1}. [We can do this if we are dealing with an invertible transformation; i.e., a transformation in which we can get $\mathbf{y}$ from $\mathbf{x}$ as $\mathbf{y} = S^{-1}\mathbf{x}$.] We obtain

$$\frac{d}{dt}\mathbf{y} = S^{-1}AS\mathbf{y}.$$

Noting that

$$\frac{d}{dt}\mathbf{y} = \Lambda\mathbf{y},$$

we have

$$\Lambda = S^{-1}AS. \tag{3.51}$$

The expression $S^{-1}AS$ is called a *similarity transformation* of matrix A. So, in order to uncouple the system, we seek a similarity transformation that results in a diagonal matrix. This process is called the *diagonalization* of matrix A. We do not know S, nor do we know Λ. We can rewrite this equation as

$$AS = S\Lambda.$$

We can solve this equation if S is *real symmetric*, i.e, $S^T = S$. [In the case of complex matrices, we need the matrix to be Hermitian, $\bar{S}^T = S$ where the bar denotes complex conjugation.]

We first show that $S\Lambda = \Lambda S$. We look at the ijth component of $S\Lambda$ and rearrange the terms in the matrix product.

$$\begin{aligned}(S\Lambda)_{ij} &= \sum_{k=1}^{n} S_{ik}\Lambda_{kj} \\ &= \sum_{k=1}^{n} S_{ik}\lambda_j I_{kj} \\ &= \sum_{k=1}^{n} \lambda_j I_{jk} S_{ki}^T \\ &= \sum_{k=1}^{n} \Lambda_{jk} S_{ki} \\ &= (\Lambda S)_{ij} \end{aligned} \tag{3.52}$$

This result leads us to the fact that S satisfies the equation

$$AS = \Lambda S.$$

Therefore, one has that the columns of S (denoted $\mathbf{v}$) satisfy an equation of the form

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{3.53}$$

This is an equation for vectors $\mathbf{v}$ and numbers λ given matrix A. It is called an *eigenvalue problem.* The vectors are called *eigenvectors* and the numbers, λ, are called *eigenvalues.* In principle, we can solve the eigenvalue problem and this will lead us to solutions of the uncoupled system of differential equations.

3.4.2 Example of an Eigenvalue Problem

We will determine the eigenvalues and eigenvectors for

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

In order to find the eigenvalues and eigenvectors of this equation, we need to solve

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{3.54}$$

Let $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Then the eigenvalue problem can be written out. We have that

$$\begin{aligned} A\mathbf{v} &= \lambda\mathbf{v} \\ \begin{pmatrix} 1 & -2 \\ -3 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= \lambda\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ \begin{pmatrix} v_1 - 2v_2 \\ -3v_1 + 2v_2 \end{pmatrix} &= \begin{pmatrix} \lambda v_1 \\ \lambda v_2 \end{pmatrix}. \end{aligned} \tag{3.55}$$

So, we see that our system becomes

$$\begin{aligned} v_1 - 2v_2 &= \lambda v_1, \\ -3v_1 + 2v_2 &= \lambda v_2. \end{aligned} \tag{3.56}$$

This can be rewritten as

$$\begin{aligned} (1-\lambda)v_1 - 2v_2 &= 0, \\ -3v_1 + (2-\lambda)v_2 &= 0. \end{aligned} \tag{3.57}$$

This is a homogeneous system. We can try to solve it using elimination, as we had done earlier when deriving Cramer's Rule. We find that multiplying the first equation by $2-\lambda$, the second by 2 and adding, we get

$$[(1-\lambda)(2-\lambda) - 6]v_1 = 0.$$

If the factor in the brackets is not zero, we obtain $v_1 = 0$. Inserting this into the system gives $v_2 = 0$ as well. Thus, we find $\mathbf{v}$ is the zero vector. However, this does not get us anywhere. We could have guessed this solution. This simple solution is the solution of all eigenvalue problems and is called the trivial solution. When solving eigenvalue problems, we only look for nontrivial solutions!

So, we have to stipulate that the factor in the brackets is zero. This means that v_1 is still unknown. This situation will always occur for eigenvalue problems. The general eigenvalue problem can be written as

$$A\mathbf{v} - \lambda\mathbf{v} = 0,$$

or by inserting the identity matrix,

$$A\mathbf{v} - \lambda I\mathbf{v} = 0.$$

Finally, we see that we always get a homogeneous system,

$$(A - \lambda I)\mathbf{v} = 0.$$

The factor that has to be zero can be seen now as the determinant of this system. Thus, we require

$$\det(A - \lambda I) = 0. \tag{3.58}$$

We write out this condition for the example at hand. We have that

$$\begin{vmatrix} 1-\lambda & -2 \\ -3 & 2-\lambda \end{vmatrix} = 0.$$

This will always be the starting point in solving eigenvalue problems. Note that the matrix is A with λ's subtracted from the diagonal elements.

Computing the determinant, we have

$$(1-\lambda)(2-\lambda) - 6 = 0,$$

or

$$\lambda^2 - 3\lambda - 4 = 0.$$

We therefore have obtained a condition on the eigenvalues! It is a quadratic and we can factor it:

$$(\lambda - 4)(\lambda + 1) = 0.$$

So, our eigenvalues are $\lambda = 4, -1$.

The second step is to find the eigenvectors. We have to do this for each eigenvalue. We first insert $\lambda = 4$ into our system:

$$\begin{aligned} -3v_1 - 2v_2 &= 0, \\ -3v_1 - 2v_2 &= 0. \end{aligned} \tag{3.59}$$

Note that these equations are the same. So, we have one equation in two unknowns. We will not get a unique solution. This is typical of eigenvalue problems. We can pick anything we want for v_2 and then determine v_1. For example, $v_2 - 1$ gives $v_1 = -2/3$. A nicer solution would be $v_2 = 3$ and $v_1 = -2$. These vectors are different, but they point in the same direction in the $v_1 v_2$ plane.

For $\lambda = -1$, the system becomes

$$\begin{aligned} 2v_1 - 2v_2 &= 0, \\ -3v_1 + 3v_2 &= 0. \end{aligned} \tag{3.60}$$

While these equations do not at first look the same, we can divide out the constants and see that once again we get the same equation,

$$v_1 = v_2.$$

Picking $v_2 = 1$, we get $v_1 = 1$.

In summary, the solution to our eigenvalue problem is

$$\lambda = 4, \quad \mathbf{v} = \begin{pmatrix} -2 \\ 3 \end{pmatrix}$$

$$\lambda = -1, \quad \mathbf{v} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$$

3.4.3 Eigenvalue Problems - A Summary

In the last subsection we were introduced to eigenvalue problems as a way to obtain a solution to a coupled system of linear differential equations. Eigenvalue problems appear in many contexts in physical applications. In this section we will summarize the method of solution of eigenvalue problems based upon our discussion in the last section. In the next subsection we will look at another problem that is a bit more geometric and will give us more insight into the process of diagonalization. We will return to our coupled system in a later section and provide more examples of solving eigenvalue problems.

We seek *nontrivial solutions* to the eigenvalue problem

$$A\mathbf{v} = \lambda\mathbf{v}. \tag{3.61}$$

We note that $\mathbf{v} = \mathbf{0}$ is an obvious solution. Furthermore, it does not lead to anything useful. So, it is a trivial solution. Typically, we are given the matrix A and have to determine the eigenvalues, λ, and the associated eigenvectors, $\mathbf{v}$, satisfying the above eigenvalue problem. Later in the course we will explore other types of eigenvalue problems.

For now we begin to solve the eigenvalue problem for $\mathbf{v} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix}$. Inserting this into Equation (3.61), we obtain the homogeneous algebraic system

$$\begin{aligned} (a-\lambda)v_1 + bv_2 &= 0, \\ cv_1 + (d-\lambda)v_2 &= 0. \end{aligned} \tag{3.62}$$

The solution of such a system would be unique if the determinant of the system is not zero. However, this would give the trivial solution $v_1 = 0$, $v_2 = 0$. To get a nontrivial solution, we need to force the determinant to be zero. This yields the eigenvalue equation

$$0 = \begin{vmatrix} a-\lambda & b \\ c & d-\lambda \end{vmatrix} = (a-\lambda)(d-\lambda) - bc.$$

This is a quadratic equation for the eigenvalues that would lead to nontrivial solutions. If we expand the right side of the equation, we find that

$$\lambda^2 - (a+d)\lambda + ad - bc = 0.$$

This is the same equation as the characteristic equation (3.90) for the general constant coefficient differential equation considered in the first chapter. Thus, the eigenvalues correspond to the solutions of the characteristic polynomial for the system.

Once we find the eigenvalues, then there are possibly an infinite number solutions to the algebraic system. We will see this in the examples.

The method for solving eigenvalue problems, as you have seen, consists of just a few simple steps. We list these steps as follows:

a) Write the coefficient matrix;

b) Find the eigenvalues from the equation $\det(A - \lambda I) = 0$; and,

c) Solve the linear system $(A - \lambda I)\mathbf{v} = 0$ for each λ.

3.4.4 Rotations of Conics

You may have seen the general form for the equation of a conic in Cartesian coordinates in your calculus class. It is given by

$$Ax^2 + 2Bxy + Cy^2 + Ex + Fy = D. \tag{3.63}$$

This equation can describe a variety of conics (ellipses, hyperbolae and parabolae) depending on the constants. The E and F terms result from a translation of the origin and the B term is the result of a rotation of the coordinate system. We leave it to the reader to show that coordinate translations can be made to eliminate the linear terms. So, we will set $E = F = 0$ in our discussion and only consider quadratic equations of the form

$$Ax^2 + 2Bxy + Cy^2 = D.$$

If $B = 0$, then the resulting equation could be an equation for the standard ellipse or hyperbola with center at the origin. In the case of an ellipse, the semimajor and semiminor axes lie along the coordinate axes. However, you could rotate the ellipse and that would introduce a B term, as we will see.

This conic equation can be written in matrix form. We note that

$$\begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} A & B \\ B & C \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = Ax^2 + 2Bxy + Cy^2.$$

In short hand matrix form, we thus have for our equation

$$\mathbf{x}^T Q \mathbf{x} = D,$$

where Q is the matrix of coefficients A, B, and C.

We want to determine the transformation that puts this conic into a coordinate system in which there is no B term. Our goal is to obtain an equation of the form

$$A'x'^2 + C'y'^2 = D'$$

in the new coordinates $\mathbf{y}^T = (x', y')$. The matrix form of this equation is given as

$$\mathbf{y}^T \begin{pmatrix} A' & 0 \\ 0 & C' \end{pmatrix} \mathbf{y} = D'.$$

We will denote the diagonal matrix by Λ.

So, we let

$$\mathbf{x} = R\mathbf{y},$$

where R is a rotation matrix. Inserting this transformation into our equation we find that

$$\begin{aligned} \mathbf{x}^T Q\mathbf{x} &= (R\mathbf{y})^T QR\mathbf{y} \\ &= \mathbf{y}^T (R^T QR)\mathbf{y}. \end{aligned} \tag{3.64}$$

Comparing this result to to desired form, we have

$$\Lambda = R^T QR. \tag{3.65}$$

Recalling that the rotation matrix is an orthogonal matrix, $R^T = R^{-1}$, we have

$$\Lambda = R^{-1} QR. \tag{3.66}$$

Thus, the problem reduces to that of trying to diagonalize the matrix Q. The eigenvalues of Q will lead to the constants in the rotated equation and the eigenvectors, as we will see, will give the directions of the principal axes (the semimajor and semiminor axes). We will first show this in an example.

Example Determine the principle axes of the ellipse given by

$$13x^2 - 10xy + 13y^2 - 72 = 0.$$

A plot of this conic in Figure 3.5 shows that it is an ellipse. However, we might not know this without plotting it. (Actually, there are some conditions on the coefficients that do allow us to determine the conic. But you may not know this yet.) If the equation were in standard form, we could identify its general shape. So, we will use the method outlined above to find a coordinate system in which the ellipse appears in standard form.

The coefficient matrix for this equation is given by

$$Q = \begin{pmatrix} 13 & -5 \\ -5 & 13 \end{pmatrix}. \tag{3.67}$$

We seek a solution to the eigenvalue problem: $Q\mathbf{v} = \lambda\mathbf{v}$. Recall, the first step is to get the eigenvalue equation from $\det(Q - \lambda I) = 0$. For this

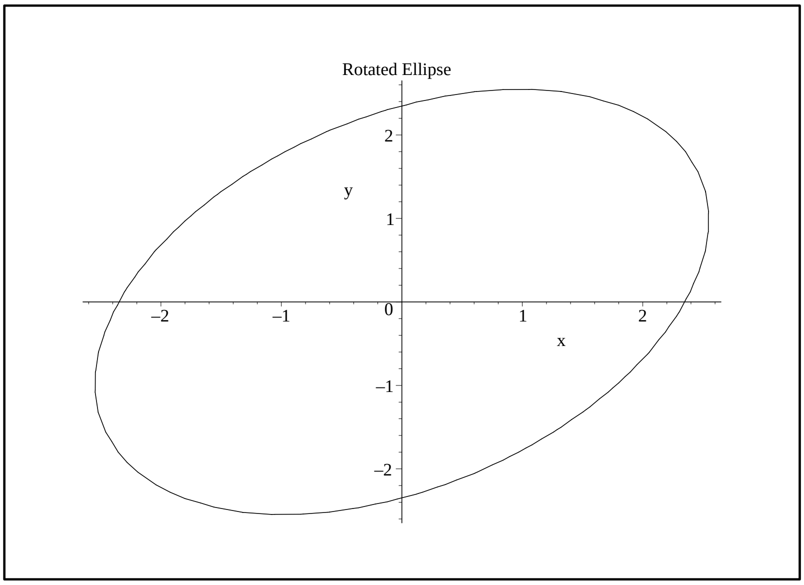

Figure 3.5: Plot of the ellipse given by $13x^2 - 10xy + 13y^2 - 72 = 0$.

problem we have

$$\begin{vmatrix} 13-\lambda & -5 \\ -5 & 13-\lambda \end{vmatrix} = 0. \tag{3.68}$$

So, we have to solve

$$(13-\lambda)^2 - 25 = 0.$$

This is easily solved by taking square roots to get

$$\lambda - 13 = \pm 5,$$

or

$$\lambda = 13 \pm 5 = 18, 8.$$

Thus, the equation in the new system is

$$8x'^2 + 18y'^2 = 72.$$

Dividing out the 72 puts this into the standard form

$$\frac{x'^2}{9} + \frac{y'^2}{4} = 1.$$

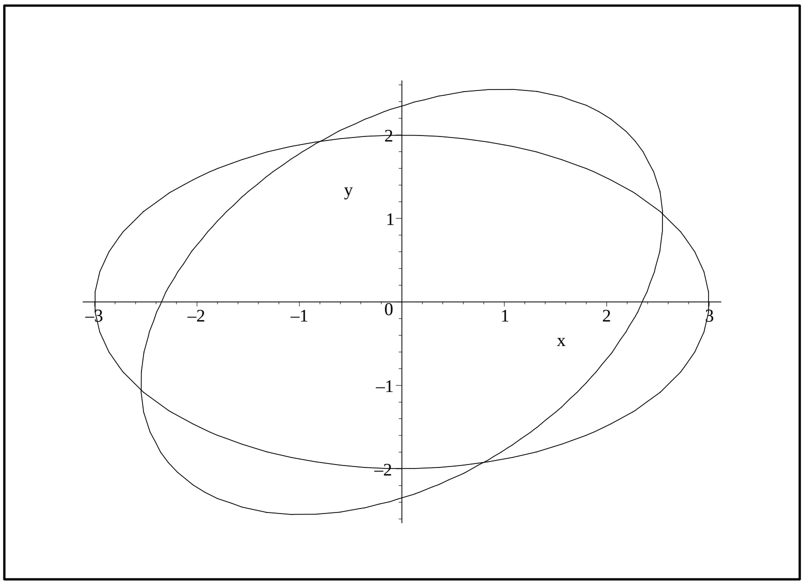

Figure 3.6: Plot of the ellipse given by $13x^2 - 10xy + 13y^2 - 72 = 0$ and the ellipse $\frac{x'^2}{9} + \frac{y'^2}{4} = 1$ showing that the first ellipse is a rotated version of the second ellipse.

Now we can identify the ellipse in the new system. We show the two ellipses in Figure 3.6. We note that the given ellipse is the new one rotated by some angle, which we still need to determine.

Next, we seek the eigenvectors corresponding to each eigenvalue.

Eigenvalue 1: $\lambda = 8$

We insert the eigenvalue into the equation $(Q - \lambda I)\mathbf{v} = 0$. The system for the unknown eigenvector is

$$\begin{pmatrix} 13-8 & -5 \\ -5 & 13-8 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0. \tag{3.69}$$

The first equation is

$$5v_1 - 5v_2 = 0, \tag{3.70}$$

or $v_1 = v_2$. Thus, we can choose our eigenvector to be

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Eigenvalue 2: $\lambda = 18$

In the same way, we insert the eigenvalue into the equation $(Q - \lambda I)\mathbf{v} = 0$ and obtain

$$\begin{pmatrix} 13-18 & -5 \\ -5 & 13-18 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 0. \tag{3.71}$$

The first equation is

$$-5v_1 - 5v_2 = 0, \tag{3.72}$$

or $v_1 = -v_2$. Thus, we can choose our eigenvector to be

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix}.$$

In Figure 3.7 we superimpose the eigenvectors on our original ellipse. We see that the eigenvectors point in directions along the semimajor and semiminor axes and indicate the angle of rotation. Eigenvector one is at a 45^o angle. Thus, our ellipse is a rotated version of one in standard position. Or, we could define new axes that are at 45^o to the standard axes and then the ellipse would take the standard form in the new coordinate system.

A general rotation of any conic can be performed. Consider the general equation:

$$Ax^2 + 2Bxy + Cy^2 + Ex + Fy = D. \tag{3.73}$$

We would like to find a rotation that puts it in the form

$$\lambda_1 x'^2 + \lambda_2 y'^2 + E'x' + F'y' = D. \tag{3.74}$$

We use the rotation matrix

$$\hat{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

and define $\mathbf{x}' = \hat{R}_\theta^T \mathbf{x}$, or $\mathbf{x} = R_\theta \mathbf{x}'$.

The general equation can be written in matrix form:

$$\mathbf{x}^T Q \mathbf{x} + \mathbf{f}\mathbf{x} = D, \tag{3.75}$$

where Q is the usual matrix of coefficients A, B, and C and $\mathbf{f} = (E, F)$. Transforming this equation gives

$$\mathbf{x}'^T R_\theta^{-1} Q R_\theta \mathbf{x}' + \mathbf{f} R_\theta \mathbf{x}' = D. \tag{3.76}$$

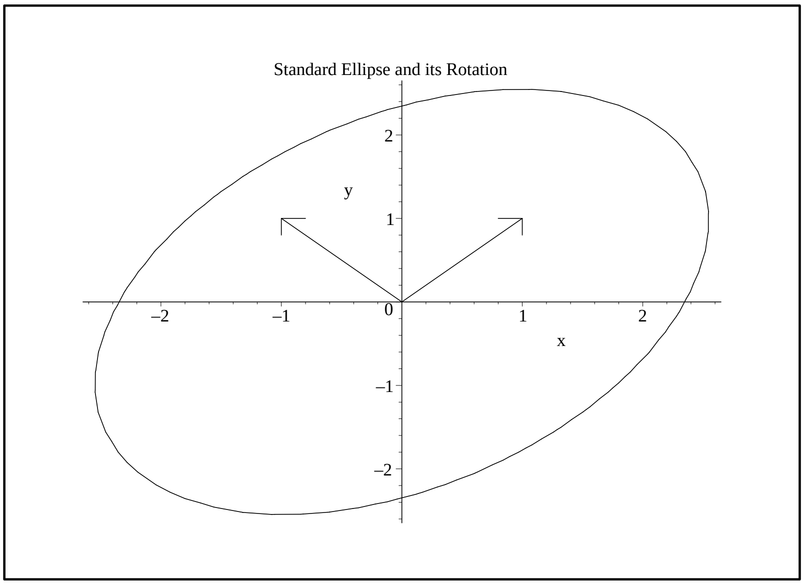

Figure 3.7: Plot of the ellipse given by $13x^2 - 10xy + 13y^2 - 72 = 0$ and the eigenvectors. Note that they are along the semimajor and semiminor axes and indicate the angle of rotation.

The resulting equation is of the form

$$A'x'^2 + 2B'x'y' + C'y'^2 + E'x' + F'y' = D, \tag{3.77}$$

where

$$B' = 2(C - A)\sin\theta\cos\theta + 2B(2\cos\theta^2 - 1). \tag{3.78}$$

(We only need B' for this discussion). If we want the nonrotated form, then we seek an angle θ such that $B' = 0$. Noting that $2\sin\theta\cos\theta = \sin 2\theta$ and $2\cos\theta^2 - 1 = \cos 2\theta$, this gives

$$\tan(2\theta) = \frac{A - C}{B}. \tag{3.79}$$

So, in our previous example, with $A = C = 13$ and $B = -5$, we have $\tan(2\theta) = \infty$. Thus, $2\theta = \pi/2$, or $\theta = \pi/4$.

Finally, we had noted that knowing the coefficients in the general quadratic is enough to determine the type of conic represented without doing any plotting. This is based on the fact that the determinant of the coefficient matrix is invariant under rotation. We see this from the equation for diagonalization

$$\begin{aligned}\det(\Lambda) &= \det(R_\theta^{-1} Q R_\theta) \\ &= \det(R_\theta^{-1}) \det(Q) \det(R_\theta) \\ &= \det(R_\theta^{-1} R_\theta) \det(Q) \\ &= \det(Q). \end{aligned} \tag{3.80}$$

Therefore, we have

$$\lambda_1 \lambda_2 = AC - B^2.$$

Looking at Equation (3.74), we have three cases:

1. Ellipse $\lambda_1\lambda_2 > 0$ or $B^2 - AC < 0$.
2. Hyperbola $\lambda_1\lambda_2 < 0$ or $B^2 - AC > 0$.
3. Parabola $\lambda_1\lambda_2 = 0$ or $B^2 - AC = 0$. and one eigenvalue is nonzero. Otherwise the equation degenerates to a linear equation.

Example $xy = 6$.

As a final example, we consider this simple equation. We can see that this is a rotated hyperbola by plotting $y = 6/x$. A plot is shown in Figure 3.8.

The coefficient matrix for this equation is given by

$$A = \begin{pmatrix} 0 & -0.5 \\ 0.5 & 0 \end{pmatrix}. \tag{3.81}$$

The eigenvalue equation is

$$\begin{vmatrix} -\lambda & -0.5 \\ -0.5 & -\lambda \end{vmatrix} = 0. \tag{3.82}$$

Thus,

$$\lambda^2 - 0.25 = 0,$$

or $\lambda = \pm 0.5$.

Once again, $\tan(2\theta) = \infty$, so the new system is at 45^o to the old. The equation in new coordinates is $0.5x^2 + (-0.5)y^2 = 6$, or $x^2 - y^2 = 12$. A plot is shown in Figure 3.9.

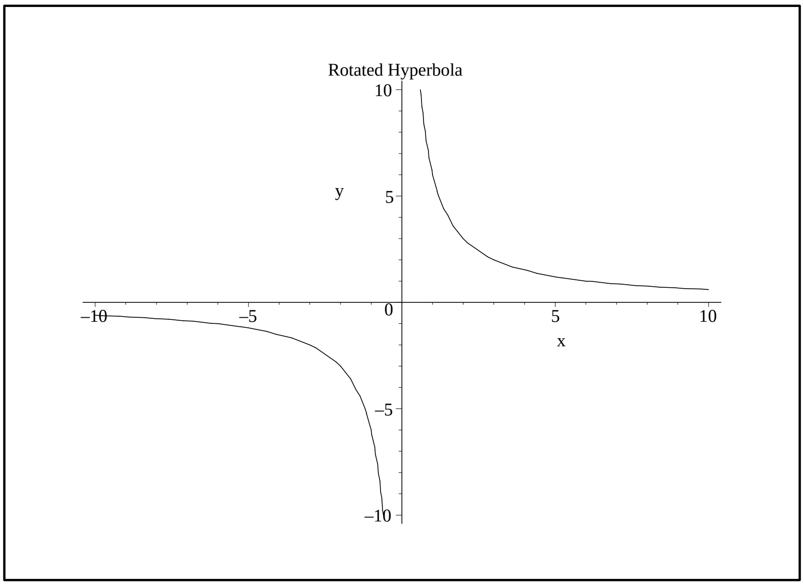

Figure 3.8: Plot of the hyperbola given by $xy = 6$.

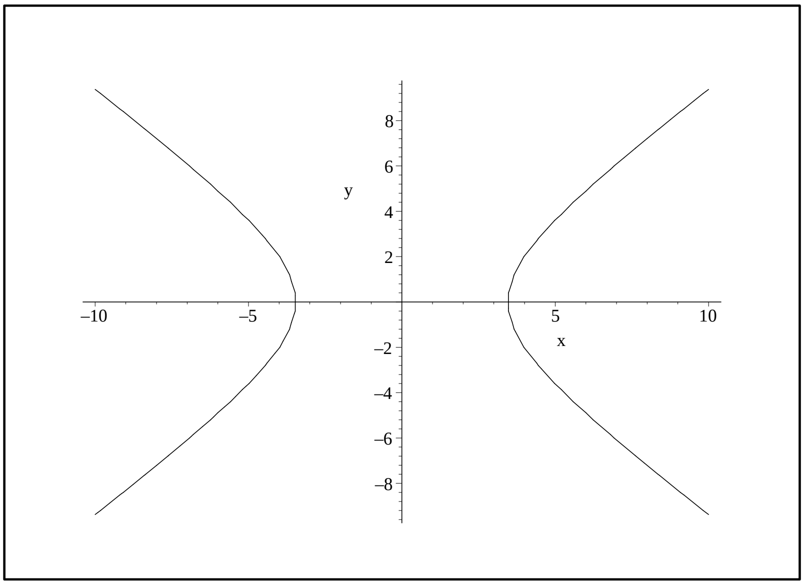

Figure 3.9: Plot of the rotated hyperbola given by $x^2 - y^2 = 12$.

3.5 A Return to Coupled Systems

We now return to examples of solving a coupled system of equations. We will review some theory of linear systems with constant coefficients. While the general techniques have already been covered, we present a bit more detail for the interested reader. We also show a few examples.

A general form for first order systems in the plane is given by a system of two equations for unknowns $x(t)$ and $y(t)$:

$$\begin{aligned} x'(t) &= P(x,y,t) \\ y'(t) &= Q(x,y,t). \end{aligned} \tag{3.83}$$

An *autonomous* system is one in which there is no explicit time dependence:

$$\begin{aligned} x'(t) &= P(x,y) \\ y'(t) &= Q(x,y). \end{aligned} \tag{3.84}$$

Otherwise the system is called *nonautonomous*.

A *linear system* takes the form

$$\begin{aligned} x' &= a(t)x + b(t)y + e(t) \\ y' &= c(t)x + d(t)y + f(t). \end{aligned} \tag{3.85}$$

A *homogeneous* linear system results when $e(t) = 0$ and $f(t) = 0$.

A *linear, constant coefficient system of differential equations* is given by

$$\begin{aligned} x' &= ax + by + e \\ y' &= cx + dy + f. \end{aligned} \tag{3.86}$$

We will focus on linear, homogeneous system of constant coefficient first order differential equations:

$$\begin{aligned} x' &= ax + by \\ y' &= cx + dy. \end{aligned} \tag{3.87}$$

As we will see later, such systems can result from a simple translation of the unknown functions. These equations are said to be coupled if either $b \neq 0$ or $c \neq 0$.

We begin by noting that the system (3.87) can be rewritten as a second order constant coefficient ordinary differential equation, which we already know how to solve. We differentiate the first equation in the system and systematically replace occurrences of y and y', since we also know from the first equation that $y = \frac{1}{b}(x' - ax)$. Thus, we have

$$\begin{aligned} x'' &= ax' + by' \\ &= ax' + b(cx + dy) \\ &= ax' + bcx + d(x' - ax). \end{aligned} \tag{3.88}$$

Therefore, we have

$$x'' - (a + d)x' + (ad - bc)x = 0. \tag{3.89}$$

This is a linear, homogeneous, constant coefficient second order ordinary differential equation. We know that we can solve this by first looking at the roots of the equation

$$r^2 - (a + d)r + ad - bc = 0 \tag{3.90}$$

and writing down the appropriate general solution for $x(t)$. Then we find $y(x) = \frac{1}{b}(x' - ax)$. We now demonstrate this for a specific example.

Example

$$\begin{aligned} x' &= -x + 6y \\ y' &= x - 2y. \end{aligned} \tag{3.91}$$

Carrying out the above steps, we have that $x'' + 3x' - 4x = 0$. This has a characteristic equation of $r^2 + 3r - 4 = 0$. The roots of this equation are $r = 1, -4$. Therefore, $x(t) = c_1 e^t + c_2 e^{-4t}$. But, we still need $y(t)$. From the first equation of the system we have

$$y(t) = \frac{1}{6}(x' + x) = \frac{1}{6}(2c_1 e^t - 3c_2 e^{-4t}).$$

Thus, the solution to our system is

$$\begin{aligned} x(t) &= c_1 e^t + c_2 e^{-4t}, \\ y(t) &= \frac{1}{3}c_1 e^t - \frac{1}{2}c_2 e^{-4t}. \end{aligned} \tag{3.92}$$

Sometimes one needs initial conditions. For these systems we would specify conditions like $x(0) = x_0$ and $y(0) = y_0$. These would allow the determination of the arbitrary constants as before.

We will next recast our system in matrix form and present a different analysis, which can easily be extended to systems of first order differential equations of more than two unknowns.

We start with the usual system in Equation (3.87). Let the unknowns be represented by the vector

$$\mathbf{x}(t) = \begin{pmatrix} x(t) \\ y(t) \end{pmatrix}.$$

Then we have that

$$\mathbf{x}' = \begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{pmatrix} ax + by \\ cx + dy \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \equiv A\mathbf{x}.$$

Here we have introduced the *coefficient matrix* A. This is a first order vector equation, $\mathbf{x}' = A\mathbf{x}$. Formerly, we can write the solution as $\mathbf{x} = \mathbf{x}_0 e^{At}$. Later, we will make some sense out of the exponential of a matrix.

We would like to investigate the solution of our system. Our investigations will lead to new techniques for solving linear systems using matrix methods.

We begin by recalling the solution to the specific problem (3.91). We obtained the solution to this system as

$$\begin{aligned} x(t) &= c_1 e^t + c_2 e^{-4t}, \\ y(t) &= \frac{1}{3} c_1 e^t - \frac{1}{2} c_2 e^{-4t}. \end{aligned} \tag{3.93}$$

This can be rewritten using matrix operations. Namely, we first write the solution in vector form.

$$\begin{aligned} \mathbf{x} &= \begin{pmatrix} x(t) \\ y(t) \end{pmatrix} = \begin{pmatrix} c_1 e^t + c_2 e^{-4t} \\ \frac{1}{3} c_1 e^t - \frac{1}{2} c_2 e^{-4t} \end{pmatrix} \\ &= \begin{pmatrix} c_1 e^t \\ \frac{1}{3} c_1 e^t \end{pmatrix} + \begin{pmatrix} c_2 e^{-4t} \\ -\frac{1}{2} c_2 e^{-4t} \end{pmatrix} \\ &= c_1 \begin{pmatrix} 1 \\ \frac{1}{3} \end{pmatrix} e^t + c_2 \begin{pmatrix} 1 \\ -\frac{1}{2} \end{pmatrix} e^{-4t}. \end{aligned} \tag{3.94}$$

We see that our solution is in the form of a linear combination of vectors of the form

$$\mathbf{x} = \mathbf{v}e^{\lambda t}$$

with $\mathbf{v}$ a constant vector and λ a constant number. This is similar to how we began to find solutions to second order constant coefficient equations. So, for the general problem (3.5) we insert this guess. Thus,

$$\begin{aligned} \mathbf{x}' &= A\mathbf{x} \Rightarrow \\ \lambda\mathbf{v}e^{\lambda t} &= A\mathbf{v}e^{\lambda t}. \end{aligned} \qquad (3.95)$$

For this to be true for all t, we then have that

$$A\mathbf{v} = \lambda\mathbf{v}. \qquad (3.96)$$

This is an eigenvalue problem. A is a 2×2 matrix for our problem, but could easily be generalized to a system of n first order differential equations. We will confine our remarks for now to planar systems. However, we need to recall how to solve eigenvalue problems and then see how solutions of eigenvalue problems can be used to obtain solutions to our systems of differential equations.

Often we are only interested in *equilibrium solutions.* For equilibrium solutions the system does not change in time. Therefore, we consider $x' = 0$ and $y' = 0$. Of course, this can only happen for constant solutions. Let x_0 and y_0 be equilibrium solutions. Then, we have

$$\begin{aligned} 0 &= ax_0 + by_0, \\ 0 &= cx_0 + dy_0. \end{aligned} \qquad (3.97)$$

This is a linear system of homogeneous algebraic equations. One only has a unique solution when the determinant of the system is not zero, $ad - bc \neq 0$. In this case, we only have the origin as a solution; i.e., $(x_0, y_0) = (0, 0)$. However, if $ad - bc = 0$, then there are an infinite number of solutions. Studies of equilibrium solutions and their stability occur more often in systems that do not readily yield to analytic solutions. Such is the case for many nonlinear systems. Such systems are the basis of research in nonlinear dynamics and chaos.

3.6 Solving Constant Coefficient Systems in 2D

Before proceeding to examples, we first indicate the types of solutions that could result from the solution of a homogeneous, constant coefficient system of first order differential equations.

We begin with the linear system of differential equations in matrix form.

$$\frac{d\mathbf{x}}{dt} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \mathbf{x} = A\mathbf{x}. \tag{3.98}$$

The type of behavior depends upon the eigenvalues of matrix A. The procedure is to determine the eigenvalues and eigenvectors and use them to construct the general solution.

If you have an initial condition,$\mathbf{x}(t_0) = \mathbf{x}_0$, you can determine your two arbitrary constants in the general solution in order to obtain the particular solution. Thus, if $\mathbf{x}_1(t)$ and $\mathbf{x}_2(t)$ are two linearly independent solutions, then the general solution is given as $\mathbf{x}(t) = c_1\mathbf{x}_1(t) + c_2\mathbf{x}_2(t)$. Then, setting $t = 0$, you get two linear equations for c_1 and $c_2 : c_1\mathbf{x}_1(0) + c_2\mathbf{x}_2(0) = \mathbf{x}_0$. We will look at a cleaner technique later in our discussion.

The major work is in finding the linearly independent solutions. This depends upon the different types of eigenvalues that you obtain from solving the eigenvalue equation, $\det(\mathbf{x} - \lambda\mathbf{I}) = 0$. The nature of these roots indicate the form of the general solution.

1. **Case I: Two real, distinct roots.**

   Solve the eigenvalue problem $A\mathbf{v} = \lambda\mathbf{v}$ for each eigenvalue obtaining two eigenvectors $\mathbf{v}_1, \mathbf{v}_2$. Then write the general solution as a linear combination $\mathbf{x}(t) = c_1e^{\lambda_1 t}\mathbf{v}_1 + c_2e^{\lambda_2 t}\mathbf{v}_2$

2. **Case II: One Repeated Root**

   Solve the eigenvalue problem $A\mathbf{v} = \lambda\mathbf{v}$ for one eigenvalue λ, obtaining the first eigenvector $\mathbf{v}_1$. One then needs a second linearly independent solution. This is obtained by solving the nonhomogeneous problem $A\mathbf{v}_2 - \lambda\mathbf{v}_2 = \mathbf{v}_1$ for $\mathbf{v}_2$.

   The general solution is then given by $\mathbf{x}(t) = c_1e^{\lambda t}\mathbf{v}_1 + c_2e^{\lambda t}(\mathbf{v}_2 + t\mathbf{v}_1)$.

3. **Case III: Two complex conjugate roots.**

   Solve the eigenvalue problem $A\mathbf{x} = \lambda\mathbf{x}$ for one eigenvalue, $\lambda = \alpha + i\beta$, obtaining one eigenvector $\mathbf{v}$. Note that this eigenvector may have complex entries. Thus, one can write the vector $\mathbf{y}(t) = e^{\lambda t}\mathbf{v} = e^{\alpha t}(\cos\beta t + i\sin\beta t)\mathbf{v}$. Now, construct two linearly independent solutions to the problem using the real and imaginary parts of $\mathbf{y}(t) : \mathbf{y}_1(t) = \mathrm{Re}(\mathbf{y}(t))$ and $\mathbf{y}_2(t) = \mathrm{Im}(\mathbf{y}(t))$. Then the general solution can be written as $\mathbf{x}(t) = c_1\mathbf{y}_1(t) + c_2\mathbf{y}_2(t)$.

The construction of the general solution in Case I is straight forward. However, the other two cases need a little explanation.

We first look at case III. Note that since the original system of equations does not have any i's, then we would expect real solutions. So, we look at the real and imaginary parts of the complex solution. We have that the complex solution satisfies the equation

$$\frac{d}{dt}[\mathrm{Re}(\mathbf{y}(t)) + i\mathrm{Im}(\mathbf{y}(t))] = A[\mathrm{Re}(\mathbf{y}(t)) + i\mathrm{Im}(\mathbf{y}(t))].$$

Differentiating the sum and splitting the real and imaginary parts of the equation, gives

$$\frac{d}{dt}\mathrm{Re}(\mathbf{y}(t)) + i\frac{d}{dt}\mathrm{Im}(\mathbf{y}(t)) = A[\mathrm{Re}(\mathbf{y}(t))] + iA[\mathrm{Im}(\mathbf{y}(t))].$$

Setting the real and imaginary parts equal, we have

$$\frac{d}{dt}\mathrm{Re}(\mathbf{y}(t)) = A[\mathrm{Re}(\mathbf{y}(t))],$$

and

$$\frac{d}{dt}\mathrm{Im}(\mathbf{y}(t)) = A[\mathrm{Im}(\mathbf{y}(t))].$$

Therefore, the real and imaginary parts each are linearly independent solutions of the system and the general solution can be written as a linear combination of these expressions.

We now turn to Case II. Writing the system of first order equations as a second order equation for $x(t)$ with the sole solution of the characteristic equation, $\lambda = \frac{1}{2}(a + d)$, we have that the general solution takes the form

$$x(t) = (c_1 + c_2 t)e^{\lambda t}.$$

This suggests that the second linearly independent solution involves a term of the form $\mathbf{v}te^{\lambda t}$. It turns out that the guess that works is

$$\mathbf{x} = te^{\lambda t}\mathbf{v}_1 + e^{\lambda t}\mathbf{v}_2.$$

Inserting this guess into the system $\mathbf{x}' = A\mathbf{x}$ yields

$$\begin{aligned}(te^{\lambda t}\mathbf{v}_1 + e^{\lambda t}\mathbf{v}_2)' &= A\left[te^{\lambda t}\mathbf{v}_1 + e^{\lambda t}\mathbf{v}_2\right] \\ e^{\lambda t}\mathbf{v}_1 + \lambda te^{\lambda t}\mathbf{v}_1 + \lambda e^{\lambda t}\mathbf{v}_2 &= \lambda te^{\lambda t}\mathbf{v}_1 + e^{\lambda t}A\mathbf{v}_2. \end{aligned} \tag{3.99}$$

Using the eigenvalue problem and noting this is true for all t, we find that

$$\mathbf{v}_1 + \lambda\mathbf{v}_2 = +A\mathbf{v}_2. \tag{3.100}$$

Therefore, $(A - \lambda I)\mathbf{v}_2 = \mathbf{v}_1$. We know everything except for $\mathbf{v}_2$. So, we just solve for it and obtain the second linearly independent solution.

3.7 Examples of the Matrix Method

Here we will give some examples of constant coefficient systems of differential equations for the three cases mentioned in the previous section. These are also examples of solving matrix eigenvalue problems.

Example 1. $A = \begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}$.

Eigenvalues: We first determine the eigenvalues.

$$0 = \begin{vmatrix} 4-\lambda & 2 \\ 3 & 3-\lambda \end{vmatrix} \tag{3.101}$$

Therefore,

$$\begin{aligned} 0 &= (4-\lambda)(3-\lambda) - 6 \\ 0 &= \lambda^2 - 7\lambda + 6 \\ 0 &= (\lambda - 1)(\lambda - 6) \end{aligned} \tag{3.102}$$

The eigenvalues are then $\lambda = 1, 6$. This is an example of Case I.

Eigenvectors: Next we determine the eigenvectors associated with each of these eigenvalues. We have to solve the system $A\mathbf{v} = \lambda\mathbf{v}$ in each case.

$\lambda = 1$.

$$\begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{3.103}$$

$$\begin{pmatrix} 3 & 2 \\ 3 & 2 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.104}$$

This gives $3v_1 + 2v_2 = 0$. One possible solution yields an eigenvector of

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2 \\ -3 \end{pmatrix}.$$

$\lambda = 6$.

$$\begin{pmatrix} 4 & 2 \\ 3 & 3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = 6\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \tag{3.105}$$

$$\begin{pmatrix} -2 & 2 \\ 3 & -3 \end{pmatrix}\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix} \tag{3.106}$$

For this case we need to solve $-2v_1 + 2v_2 = 0$. This yields

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

General Solution: We can now construct the general solution.

$$\begin{aligned} \mathbf{x}(t) &= c_1 e^{\lambda_1 t}\mathbf{v}_1 + c_2 e^{\lambda_2 t}\mathbf{v}_2 \\ &= c_1 e^{t}\begin{pmatrix} 2 \\ -3 \end{pmatrix} + c_2 e^{6t}\begin{pmatrix} 1 \\ 1 \end{pmatrix} \\ &= \begin{pmatrix} 2c_1 e^{t} + c_2 e^{6t} \\ -3c_1 e^{t} + c_2 e^{6t} \end{pmatrix}. \end{aligned} \tag{3.107}$$

Example 2. $A = \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix}$.

Eigenvalues: Again, one solves the eigenvalue equation.

$$0 = \begin{vmatrix} 3-\lambda & -5 \\ 1 & -1-\lambda \end{vmatrix} \tag{3.108}$$

Therefore,

$$\begin{aligned} 0 &= (3-\lambda)(-1-\lambda)+5 \\ 0 &= \lambda^2 - 2\lambda + 2 \\ \lambda &= \frac{-(-2) \pm \sqrt{4-4(1)(2)}}{2} = 1 \pm i. \end{aligned} \tag{3.109}$$

The eigenvalues are then $\lambda = 1+i, 1-i$. This is an example of Case III.

Eigenvectors: In order to find the general solution, we need only find the eigenvector associated with $1+i$.

$$\begin{aligned} \begin{pmatrix} 3 & -5 \\ 1 & -1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= (1+i) \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\ \begin{pmatrix} 2-i & -5 \\ 1 & -2-i \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \end{aligned} \tag{3.110}$$

We need to solve $(2-i)v_1 - 5v_2 = 0$. Thus,

$$\begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 2+i \\ 1 \end{pmatrix}. \tag{3.111}$$

Complex Solution: In order to get the two real linearly independent solutions, we need to compute the real and imaginary parts of $\mathbf{v}e^{\lambda t}$.

$$\begin{aligned} e^{\lambda t} \begin{pmatrix} 2+i \\ 1 \end{pmatrix} &= e^{(1+i)t} \begin{pmatrix} 2+i \\ 1 \end{pmatrix} \\ &= e^t(\cos t + i \sin t) \begin{pmatrix} 2+i \\ 1 \end{pmatrix} \\ &= e^t \begin{pmatrix} (2+i)(\cos t + i \sin t) \\ \cos t + i \sin t \end{pmatrix} \end{aligned}$$

$$
\begin{aligned}
&= e^t \begin{pmatrix} (2\cos t - \sin t) + i(\cos t + 2\sin t) \\ \cos t + i \sin t \end{pmatrix} \\
&= e^t \begin{pmatrix} 2\cos t - \sin t \\ \cos t \end{pmatrix} + ie^t \begin{pmatrix} \cos t + 2\sin t \\ \sin t \end{pmatrix}.
\end{aligned}
$$

General Solution: Now we can construct the general solution.

$$
\begin{aligned}
\mathbf{x}(t) &= c_1 e^t \begin{pmatrix} 2\cos t - \sin t \\ \cos t \end{pmatrix} + c_2 e^t \begin{pmatrix} \cos t + 2\sin t \\ \sin t \end{pmatrix} \\
&= e^t \begin{pmatrix} c_1(2\cos t - \sin t) + c_2(\cos t + 2\sin t) \\ c_1 \cos t + c_2 \sin t \end{pmatrix}. \quad (3.112)
\end{aligned}
$$

Note: This can be rewritten as

$$
\mathbf{x}(t) = e^t \cos t \begin{pmatrix} 2c_1 + c_2 \\ c_1 \end{pmatrix} + e^t \sin t \begin{pmatrix} 2c_2 - c_1 \\ c_2 \end{pmatrix}.
$$

Example 3. $A = \begin{pmatrix} 7 & -1 \\ 9 & 1 \end{pmatrix}$.

Eigenvalues:

$$
0 = \begin{vmatrix} 7-\lambda & -1 \\ 9 & 1-\lambda \end{vmatrix} \quad (3.113)
$$

Therefore,

$$
\begin{aligned}
0 &= (7-\lambda)(1-\lambda) + 9 \\
0 &= \lambda^2 - 8\lambda + 16 \\
0 &= (\lambda - 4)^2. \quad (3.114)
\end{aligned}
$$

There is only one real eigenvalue, $\lambda = 4$. This is an example of Case II.

Eigenvectors: In this case we first solve for $\mathbf{v}_1$ and then get the second linearly independent vector.

$$
\begin{aligned}
\begin{pmatrix} 7 & -1 \\ 9 & 1 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= 4 \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} \\
\begin{pmatrix} 3 & -1 \\ 9 & -3 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix}. \quad (3.115)
\end{aligned}
$$

Therefore, we have

$$3v_1 - v_2 = 0, \quad \Rightarrow \quad \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 3 \end{pmatrix}.$$

Second Linearly Independent Solution:

Now we need to solve $A\mathbf{v}_2 - \lambda\mathbf{v}_2 = \mathbf{v}_1$.

$$\begin{aligned} \begin{pmatrix} 7 & -1 \\ 9 & 1 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} - 4\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= \begin{pmatrix} 1 \\ 3 \end{pmatrix} \\ \begin{pmatrix} 3 & -1 \\ 9 & -3 \end{pmatrix}\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} &= \begin{pmatrix} 1 \\ 3 \end{pmatrix}. \end{aligned} \tag{3.116}$$

We therefore need to solve the system of equations

$$\begin{aligned} 3u_1 - u_2 &= 1 \\ 9u_1 - 3u_2 &= 3. \end{aligned} \tag{3.117}$$

The solution is $\begin{pmatrix} u_1 \\ u_2 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$.

General Solution: We construct the general solution as

$$\begin{aligned} \mathbf{y}(t) &= c_1 e^{\lambda t}\mathbf{v}_1 + c_2 e^{\lambda t}(\mathbf{v}_2 + t\mathbf{v}_1). \\ &= c_1 e^{4t}\begin{pmatrix} 1 \\ 3 \end{pmatrix} + c_2 e^{4t}\left[\begin{pmatrix} 1 \\ 2 \end{pmatrix} + t\begin{pmatrix} 1 \\ 3 \end{pmatrix}\right] \\ &= e^{4t}\begin{pmatrix} c_1 + c_2(1+t) \\ 3c_1 + c_2(2+3t) \end{pmatrix}. \end{aligned} \tag{3.118}$$

3.8 Inner Product Spaces *Optional*

This chapter has been about some of the linear algebra background that is needed in undergraduate physics. We have only discussed finite dimensional vector spaces, linear transformations and their matrix representations, and solving eigenvalue problems. There is more that we could discuss and more rigor. As we progress through the course we will return to the basics and some of their generalizations. An important generalization for physics is to infinite dimensional vector spaces, in particular - function spaces. This conceptual framework is very important in areas such as quantum mechanics because this is the basis of solving the eigenvalue problems that come up there so often with the Schrödinger equation. We will also see in the next chapter that the appropriate background spaces are function spaces in which we can solve the wave and heat equations. While we do not immediately need this understanding to carry out our computations, it can later help in the overall understanding of the methods of solution of linear partial differential equations. In particular, one definitely needs to grasp these ideas in order to fully understand and appreciate quantum mechanics.

We will consider the space of functions of a certain type. They could be the space of continuous functions on [0,1], or the space of differentiably continuous functions, or the set of functions integrable from a to b. Later, we will specify the types of functions. However, you can see that there are many types of function spaces. We will further need to be able to add functions and multiply them by scalars. Thus, the set of functions and these operations will provide us with a vector space of functions.

We will also need a scalar product defined on this space of functions. There are several types of scalar products, or inner products, that we can define. For a real vector space, we define

An *inner product* $<,>$ on a real vector space V is a mapping from $V \times V$ into R such that for $u, v, w \in V$ and $\alpha \in R$ one has

1. $< v, v > \geq 0$ and $< v, v >= 0$ if and only if $v = 0$.
2. $< v, w >=< w, v >$.
3. $< \alpha v, w >= \alpha < v, w >$.
4. $< u + v, w >=< u, w > + < v, w >$.

A real vector space equipped with the above inner product leads to a real inner product space. A more general definition with the second property replaced with $< v, w >= \overline{< w, v >}$ is needed for complex inner product spaces.

For the time being, we are dealing just with real valued functions. We need an inner product appropriate for such spaces. One such definition is the following. Let $f(x)$ and $g(x)$ be functions defined on $[a, b]$. Then, we define the inner product, if the integral exists, as

$$< f, g >= \int_a^b f(x)g(x)\, dx. \tag{3.119}$$

So, we have functions spaces equipped with an inner product. Can we find a basis for the space? For an n-dimensional space we need n basis vectors. For an infinite dimensional space, how many will we need? How do we know when we have enough? We will consider the answers to these questions as we proceed through the text.

Let's assume that we have a basis of functions $\{\phi_n(x)\}_{n=1}^{\infty}$. Given a function $f(x)$, how can we go about finding the components of f in this basis? In other words, let

$$f(x) = \sum_{n=1}^{\infty} c_n \phi_n(x).$$

How do we find the c_n's? Does this remind you of the problem we had earlier?

Formally, we take the inner product of f with each ϕ_j, to find

$$\begin{aligned} < \phi_j, f > &= < \phi_j, \sum_{n=1}^{\infty} c_n \phi_n > \\ &= \sum_{n=1}^{\infty} c_n < \phi_j, \phi_n > . \end{aligned} \tag{3.120}$$

If our basis is an orthogonal basis, then we have

$$< \phi_j, \phi_n >= N_j \delta_{ij}, \tag{3.121}$$

where δ_{ij} is the Kronecker delta.

Thus, we have

$$\begin{aligned} < \phi_j, f > &= \sum_{n=1}^{\infty} c_n < \phi_j, \phi_n > \\ &= \sum_{n=1}^{\infty} c_n N_j \delta_{ij} \\ &= c_j N_j. \end{aligned} \tag{3.122}$$

So, the expansion coefficient is

$$c_j = \frac{< \phi_j, f >}{N_j} = \frac{< \phi_j, f >}{< \phi_j, \phi_j >}.$$

In our preparation for later sections, let's determine if the set of functions $\phi_n(x) = \sin nx$ for $n = 1, 2, \ldots$ is orthogonal on the interval $[= \pi, \pi]$. We need to show that $< \phi_n, \phi_m >= 0$ for $n \neq m$. Thus, we have for $n \neq m$

$$\begin{aligned} < \phi_n, \phi_m > &= \int_{-\pi}^{\pi} \sin nx \sin mx \, dx \\ &= \frac{1}{2} \int_{-\pi}^{\pi} [\cos(n-m)x - \cos(n+m)x] \, dx \\ &= \frac{1}{2} \left[\frac{\sin(n-m)x}{n-m} - \frac{\sin(n+m)x}{n+m} \right]_{-\pi}^{\pi} = 0. \end{aligned} \tag{3.123}$$

Here we have made use of a trigonometric identity for the product of two sines.

So, we have determined that the set $\phi_n(x) = \sin nx$ for $n = 1, 2, \ldots$ is an orthogonal set of functions on the interval $[= \pi, \pi]$. Just as with vectors in three dimensions, we can normalize our basis functions to arrive at an orthonormal basis. This is simply done by dividing by the *length* of the vector. Recall that the length of a vector was obtained as $v = \sqrt{\mathbf{v} \cdot \mathbf{v}}$. In the same way, we define the *norm* of our functions by

$$\|f\| = \sqrt{< f, f >}.$$

Note, there are many types of norms, but this will be sufficient for us.

For the above basis of sine functions, we want to first compute the norm of each function. Then we would like to find a new basis from this one such that each basis eigenfunction has unit length and is therefore an orthonormal basis. We first compute

$$\begin{aligned} \|\phi_n\|^2 &= \int_{-\pi}^{\pi} \sin^2 nx \, dx \\ &= \frac{1}{2} \int_{-\pi}^{\pi} [1 - \cos 2nx] \, dx \\ &= \frac{1}{2} \left[x - \frac{\sin 2nx}{2n} \right]_{-\pi}^{\pi} = \pi. \end{aligned} \tag{3.124}$$

We have found from this computation that

$$< \phi_j, \phi_n >= \pi\delta_{ij} \tag{3.125}$$

and that $\|\phi_n\| = \sqrt{\pi}$. Defining $\psi_n(x) = \frac{1}{\sqrt{\pi}}\phi_n(x)$, we have normalized the ϕ_n's and have obtained an orthonormal basis of functions on $[-\pi, \pi]$.

Expansions of functions in trigonometric bases occur often and originally resulted from the study of partial differential equations. They have been named Fourier series and will be the topic of the next chapter.