6 Linear Transformations for Neural Networks

Objectives

This chapter will continue the work of Chapter 5 in laying out the mathematical foundations for our analysis of neural networks. In Chapter 5 we reviewed vector spaces; in this chapter we investigate linear transformations as they apply to neural networks.

As we have seen in previous chapters, the multiplication of an input vector by a weight matrix is one of the key operations that is performed by neural networks. This operation is an example of a linear transformation. We want to investigate general linear transformations and determine their fundamental characteristics. The concepts covered in this chapter, such as eigenvalues, eigenvectors and change of basis, will be critical to our understanding of such key neural network topics as performance learning (including the Widrow-Hoff rule and backpropagation) and Hopfield network convergence.

Theory and Examples

Recall the Hopfield network that was discussed in Chapter 3. (See Figure 6.1.) The output of the network is updated synchronously according to the equation

$$\mathbf{a}(t+1) = satlin(\mathbf{W}\mathbf{a}(t) + \mathbf{b}) . \tag{6.1}$$

Notice that at each iteration the output of the network is again multiplied by the weight matrix $\mathbf{W}$. What is the effect of this repeated operation? Can we determine whether or not the output of the network will converge to some steady state value, go to infinity, or oscillate? In this chapter we will lay the foundation for answering these questions, along with many other questions about neural networks discussed in this book.

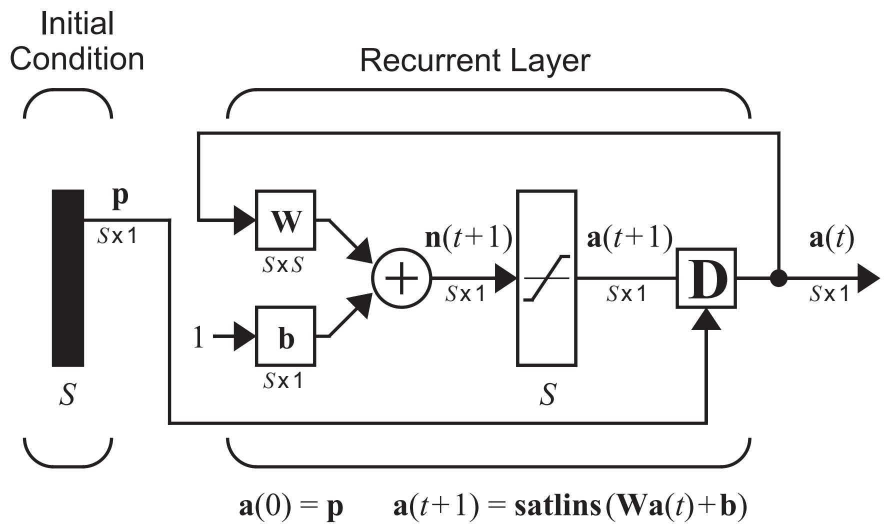


Figure 6.1 Hopfield Network

Linear Transformations

We begin with some general definitions.

Transformation

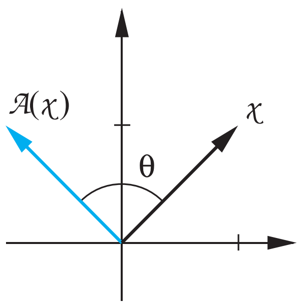
A *transformation* consists of three parts:

1. a set of elements $X = \{\chi_i\}$, called the domain,
2. a set of elements $Y = \{y_i\}$, called the range, and
3. a rule relating each $\chi_i \in X$ to an element $y_i \in Y$.

Linear Transformation

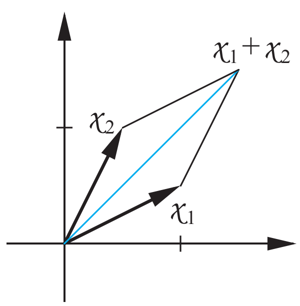
A transformation $\mathcal{A}$ is *linear* if:

1. for all $\mathbf{x}_1, \mathbf{x}_2 \in X$, $\mathcal{A}(\mathbf{x}_1 + \mathbf{x}_2) = \mathcal{A}(\mathbf{x}_1) + \mathcal{A}(\mathbf{x}_2)$,
2. for all $\mathbf{x} \in X$, $a \in R$, $\mathcal{A}(a\mathbf{x}) = a\mathcal{A}(\mathbf{x})$.



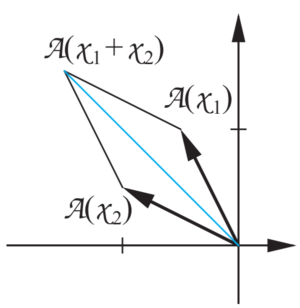
Consider, for example, the transformation obtained by rotating vectors in $\Re^2$ by an angle θ, as shown in the figure to the left. The next two figures illustrate that property 1 is satisfied for rotation. They show that if you want to rotate a sum of two vectors, you can rotate each vector first and then sum them. The fourth figure illustrates property 2. If you want to rotate a scaled vector, you can rotate it first and then scale it. Therefore rotation is a linear operation.

Matrix Representations



As we mentioned at the beginning of this chapter, matrix multiplication is an example of a linear transformation. We can also show that any linear transformation between two finite-dimensional vector spaces can be represented by a matrix (just as in the last chapter we showed that any general vector in a finite-dimensional vector space can be represented by a column of numbers). To show this we will use most of the concepts covered in the previous chapter.

Let $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n\}$ be a basis for vector space X, and let $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_m\}$ be a basis for vector space Y. This means that for any two vectors $\mathbf{x} \in X$ and $\mathbf{y} \in Y$

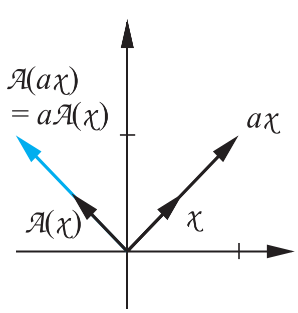


$$\mathbf{x} = \sum_{i=1}^{n} x_i \mathbf{v}_i \text{ and } \mathbf{y} = \sum_{i=1}^{m} y_i \mathbf{u}_i . \tag{6.2}$$

Let $\mathcal{A}$ be a linear transformation with domain X and range Y ($\mathcal{A}: X \to Y$). Then

$$\mathcal{A}(\mathbf{x}) = \mathbf{y} \tag{6.3}$$

can be written



$$\mathcal{A}\left(\sum_{j=1}^{n} x_j \mathbf{v}_j\right) = \sum_{i=1}^{m} y_i \mathbf{u}_i . \tag{6.4}$$

Since $\mathcal{A}$ is a linear operator, Eq. (6.4) can be written

$$\sum_{j=1}^{n} x_j \mathcal{A}(\mathbf{v}_j) = \sum_{i=1}^{m} y_i \mathbf{u}_i . \tag{6.5}$$

Since the vectors $\mathcal{A}(v_j)$ are elements of Y, they can be written as linear combinations of the basis vectors for Y:

$$\mathcal{A}(v_j) = \sum_{i=1}^{m} a_{ij} u_i . \tag{6.6}$$

(Note that the notation used for the coefficients of this expansion, a_{ij}, was not chosen by accident.) If we substitute Eq. (6.6) into Eq. (6.5) we obtain

$$\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij} u_i = \sum_{i=1}^{m} y_i u_i . \tag{6.7}$$

The order of the summations can be reversed, to produce

$$\sum_{i=1}^{m} u_i \sum_{j=1}^{n} a_{ij} x_j = \sum_{i=1}^{m} y_i u_i . \tag{6.8}$$

This equation can be rearranged, to obtain

$$\sum_{i=1}^{m} u_i \left(\sum_{j=1}^{n} a_{ij} x_j - y_i \right) = 0. \tag{6.9}$$

Recall that since the u_i form a basis set they must be independent. This means that each coefficient that multiplies u_i in Eq. (6.9) must be identically zero (see Eq. (5.4)), therefore

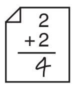
$$\sum_{j=1}^{n} a_{ij} x_j = y_i . \tag{6.10}$$

This is just matrix multiplication, as in

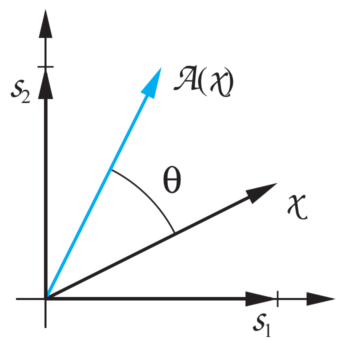
$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} \tag{6.11}$$

We can summarize these results: *For any linear transformation between two finite-dimensional vector spaces there is a matrix representation. When we multiply the matrix times the vector expansion for the domain vector χ, we obtain the vector expansion for the transformed vector y.*

Keep in mind that the matrix representation is not unique (just as the representation of a general vector by a column of numbers is not unique — see Chapter 5). If we change the basis set for the domain or for the range, the matrix representation will also change. We will use this fact to our advantage in later chapters.

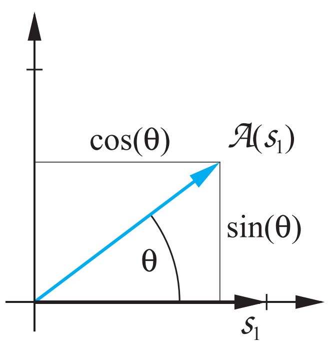


As an example of a matrix representation, consider the rotation transformation. Let's find a matrix representation for that transformation. The key step is given in Eq. (6.6). We must transform each basis vector for the domain and then expand it in terms of the basis vectors of the range. In this example the domain and the range are the same ($X = Y = \Re^2$), so to keep things simple we will use the standard basis for both ($u_i = v_i = s_i$), as shown in the adjacent figure.



The first step is to transform the first basis vector and expand the resulting transformed vector in terms of the basis vectors. If we rotate s_1 counterclockwise by the angle θ we obtain

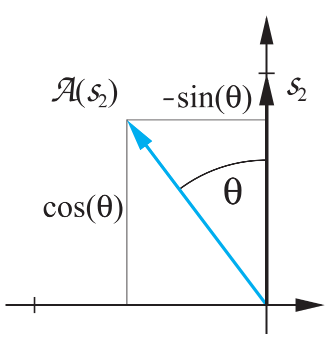
$$\mathcal{A}(s_1) = \cos(\theta)s_1 + \sin(\theta)s_2 = \sum_{i=1}^{2} a_{i1}s_i = a_{11}s_1 + a_{21}s_2, \qquad (6.12)$$



as can be seen in the middle left figure. The two coefficients in this expansion make up the first column of the matrix representation.

The next step is to transform the second basis vector. If we rotate s_2 counterclockwise by the angle θ we obtain

$$\mathcal{A}(s_2) = -\sin(\theta)s_1 + \cos(\theta)s_2 = \sum_{i=1}^{2} a_{i2}s_i = a_{12}s_1 + a_{22}s_2, \qquad (6.13)$$



as can be seen in the lower left figure. From this expansion we obtain the second column of the matrix representation. The complete matrix representation is thus given by

$$\mathbf{A} = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix}. \qquad (6.14)$$

Verify for yourself that when you multiply a vector by the matrix of Eq. (6.14), the vector is rotated by an angle θ.

In summary, to obtain the matrix representation of a transformation we use Eq. (6.6). We transform each basis vector for the domain and expand it in terms of the basis vectors of the range. The coefficients of each expansion produce one column of the matrix.

To graphically investigate the process of creating a matrix representation, use the Neural Network Design Demonstration Linear Transformations (`nnd6lt`).

Change of Basis

We notice from the previous section that the matrix representation of a linear transformation is not unique. The representation will depend on what basis sets are used for the domain and the range of the transformation. In this section we will illustrate exactly how a matrix representation changes as the basis sets are changed.

Consider a linear transformation $\mathcal{A}: X \to Y$. Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for vector space X, and let $\{u_1, u_2, \ldots, u_m\}$ be a basis for vector space Y. Therefore, any vector $\chi \in X$ can be written

$$\chi = \sum_{i=1}^{n} x_i v_i , \tag{6.15}$$

and any vector $y \in Y$ can be written

$$y = \sum_{i=1}^{m} y_i u_i . \tag{6.16}$$

So if

$$\mathcal{A}(\chi) = y \tag{6.17}$$

the matrix representation will be

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix} , \tag{6.18}$$

or

$$\mathbf{A}\mathbf{x} = \mathbf{y} . \tag{6.19}$$

Now suppose that we use different basis sets for X and Y. Let $\{t_1, t_2, \ldots, t_n\}$ be the new basis for X, and let $\{w_1, w_2, \ldots, w_m\}$ be the new basis for Y. With the new basis sets, the vector $\chi \in X$ is written

$$\chi = \sum_{i=1}^{n} x'_i t_i , \qquad (6.20)$$

and the vector $y \in Y$ is written

$$y = \sum_{i=1}^{m} y'_i w_i . \qquad (6.21)$$

This produces a new matrix representation:

$$\begin{bmatrix} a'_{11} & a'_{12} & \dots & a'_{1n} \\ a'_{21} & a'_{22} & \dots & a'_{2n} \\ \vdots & \vdots & & \vdots \\ a'_{m1} & a'_{m2} & \dots & a'_{mn} \end{bmatrix} \begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_n \end{bmatrix} = \begin{bmatrix} y'_1 \\ y'_2 \\ \vdots \\ y'_m \end{bmatrix} , \qquad (6.22)$$

or

$$\mathbf{A}'\mathbf{x}' = \mathbf{y}' . \qquad (6.23)$$

What is the relationship between $\mathbf{A}$ and $\mathbf{A}'$? To find out, we need to find the relationship between the two basis sets. First, since each t_i is an element of X, they can be expanded in terms of the original basis for X:

$$t_i = \sum_{j=1}^{n} t_{ji} v_j . \qquad (6.24)$$

Next, since each w_i is an element of Y, they can be expanded in terms of the original basis for Y:

$$w_i = \sum_{j=1}^{m} w_{ji} u_j . \qquad (6.25)$$

Therefore, the basis vectors can be written as columns of numbers:

$$\mathbf{t}_i = \begin{bmatrix} t_{1i} \\ t_{2i} \\ \vdots \\ t_{ni} \end{bmatrix} \qquad \mathbf{w}_i = \begin{bmatrix} w_{1i} \\ w_{2i} \\ \vdots \\ w_{mi} \end{bmatrix} . \qquad (6.26)$$

Define a matrix whose columns are the $\mathbf{t}_i$:

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \dots & \mathbf{t}_n \end{bmatrix} . \tag{6.27}$$

Then we can write Eq. (6.20) in matrix form:

$$\mathbf{x} = x'_1 \mathbf{t}_1 + x'_2 \mathbf{t}_2 + \cdots + x'_n \mathbf{t}_n = \mathbf{B}_t \mathbf{x}' . \tag{6.28}$$

This equation demonstrates the relationships between the two different representations for the vector χ. (Note that this is effectively the same as Eq. (5.43). You may want to revisit our discussion of reciprocal basis vectors in Chapter 5.)

Now define a matrix whose columns are the $\mathbf{w}_i$:

$$\mathbf{B}_w = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \dots & \mathbf{w}_m \end{bmatrix} . \tag{6.29}$$

This allows us to write Eq. (6.21) in matrix form,

$$\mathbf{y} = \mathbf{B}_w \mathbf{y}' , \tag{6.30}$$

which then demonstrates the relationships between the two different representations for the vector y.

Now substitute Eq. (6.28) and Eq. (6.30) into Eq. (6.19):

$$\mathbf{A} \mathbf{B}_t \mathbf{x}' = \mathbf{B}_w \mathbf{y}' . \tag{6.31}$$

If we multiply both sides of this equation by $\mathbf{B}_w^{-1}$ we obtain

$$[\mathbf{B}_w^{-1} \mathbf{A} \mathbf{B}_t] \mathbf{x}' = \mathbf{y}' . \tag{6.32}$$

Change of Basis

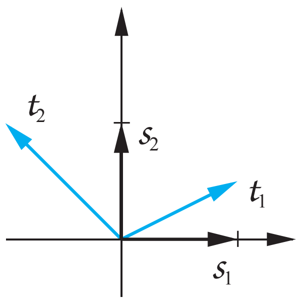
A comparison of Eq. (6.32) and Eq. (6.23) yields the following operation for a *change of basis*:

$$\mathbf{A}' = [\mathbf{B}_w^{-1} \mathbf{A} \mathbf{B}_t] . \tag{6.33}$$

Similarity Transform

This key result, which describes the relationship between any two matrix representations of a given linear transformation, is called a *similarity transform* [Brog91]. It will be of great use to us in later chapters. It turns out that with the right choice of basis vectors we can obtain a matrix representation that reveals the key characteristics of the linear transformation it represents. This will be discussed in the next section.

As an example of changing basis sets, let's revisit the vector rotation example of the previous section. In that section a matrix representation was developed using the standard basis set $\{s_1, s_2\}$. Now let's find a new representation using the basis $\{t_1, t_2\}$, which is shown in the adjacent fig-



ure. (Note that in this example the same basis set is used for both the domain and the range.)

The first step is to expand t_1 and t_2 in terms of the standard basis set, as in Eq. (6.24) and Eq. (6.25). By inspection of the adjacent figure we find:

$$t_1 = s_1 + 0.5 s_2 , \tag{6.34}$$

$$t_2 = -s_1 + s_2 . \tag{6.35}$$

Therefore we can write

$$\mathbf{t}_1 = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} \qquad \mathbf{t}_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} . \tag{6.36}$$

Now we can form the matrix

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 \\ 0.5 & 1 \end{bmatrix} , \tag{6.37}$$

and, because we are using the same basis set for both the domain and the range of the transformation,

$$\mathbf{B}_w = \mathbf{B}_t = \begin{bmatrix} 1 & -1 \\ 0.5 & 1 \end{bmatrix} . \tag{6.38}$$

We can now compute the new matrix representation from Eq. (6.33):

$$\mathbf{A}' = [\mathbf{B}_w^{-1} \mathbf{A} \mathbf{B}_t] = \begin{bmatrix} 2/3 & 2/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0.5 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 1/3 \sin\theta + \cos\theta & -4/3 \sin\theta \\ \frac{5}{6} \sin\theta & -1/3 \sin\theta + \cos\theta \end{bmatrix} . \tag{6.39}$$

Take, for example, the case where $\theta = 30°$.

$$\mathbf{A}' = \begin{bmatrix} 1.033 & -0.667 \\ 0.417 & 0.699 \end{bmatrix} , \tag{6.40}$$

and

$$\mathbf{A} = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix}. \qquad (6.41)$$

To check that these matrices are correct, let's try a test vector

$$\mathbf{x} = \begin{bmatrix} 1 \\ 0.5 \end{bmatrix}, \text{ which corresponds to } \mathbf{x}' = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \qquad (6.42)$$

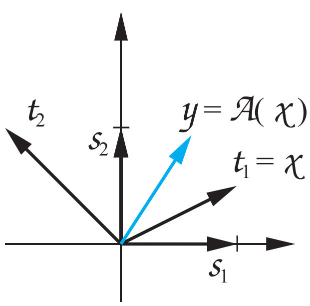
(Note that the vector represented by $\mathbf{x}$ and $\mathbf{x}'$ is t_1, a member of the second basis set.) The transformed test vector would be

$$\mathbf{y} = \mathbf{A}\mathbf{x} = \begin{bmatrix} 0.866 & -0.5 \\ 0.5 & 0.866 \end{bmatrix} \begin{bmatrix} 1 \\ 0.5 \end{bmatrix} = \begin{bmatrix} 0.616 \\ 0.933 \end{bmatrix}, \qquad (6.43)$$

which should correspond to

$$\mathbf{y}' = \mathbf{A}'\mathbf{x}' = \begin{bmatrix} 1.033 & -0.667 \\ 0.416 & 0.699 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1.033 \\ 0.416 \end{bmatrix}. \qquad (6.44)$$

How can we test to see if $\mathbf{y}'$ does correspond to $\mathbf{y}$? Both should be representations of the same vector, y, in terms of two different basis sets; $\mathbf{y}$ uses the basis $\{s_1, s_2\}$ and $\mathbf{y}'$ uses the basis $\{t_1, t_2\}$. In Chapter 5 we used the reciprocal basis vectors to transform from one representation to another (see Eq. (5.43)). Using that concept we have



$$\mathbf{y}' = \mathbf{B}^{-1}\mathbf{y} = \begin{bmatrix} 1 & -1 \\ 0.5 & 1 \end{bmatrix}^{-1} \begin{bmatrix} 0.616 \\ 0.933 \end{bmatrix} = \begin{bmatrix} 2/3 & 2/3 \\ -1/3 & 2/3 \end{bmatrix} \begin{bmatrix} 0.616 \\ 0.933 \end{bmatrix} = \begin{bmatrix} 1.033 \\ 0.416 \end{bmatrix}, \qquad (6.45)$$

which verifies our previous result. The vectors are displayed in the figure to the left. Verify graphically that the two representations, $\mathbf{y}$ and $\mathbf{y}'$, given by Eq. (6.43) and Eq. (6.44), are reasonable.

Eigenvalues and Eigenvectors

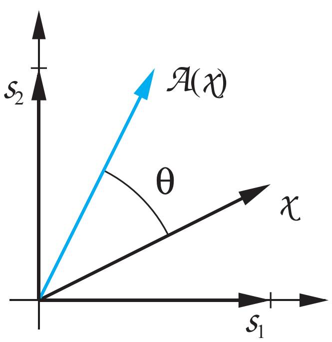
In this final section we want to discuss two key properties of linear transformations: eigenvalues and eigenvectors. Knowledge of these properties will allow us to answer some key questions about neural network performance, such as the question we posed at the beginning of this chapter, concerning the stability of Hopfield networks.

Eigenvalues
Eigenvectors

Let's first define what we mean by *eigenvalues* and *eigenvectors*. Consider a linear transformation $\mathcal{A}: X \to X$. (The domain is the same as the range.) Those vectors $z \in X$ that are not equal to zero and those scalars λ that satisfy

$$\mathcal{A}(z) = \lambda z \qquad (6.46)$$

are called eigenvectors (z) and eigenvalues (λ), respectively. Notice that the term eigenvector is a little misleading, since it is not really a vector but a vector space, since if z satisfies Eq. (6.46), then az will also satisfy it.



Therefore an eigenvector of a given transformation represents a direction, such that any vector in that direction, when transformed, will continue to point in the same direction, but will be scaled by the eigenvalue. As an example, consider again the rotation example used in the previous sections. Is there any vector that, when rotated by 30°, continues to point in the same direction? No; this is a case where there are no real eigenvalues. (If we allow complex scalars, then two eigenvalues exist, as we will see later.)

How can we compute the eigenvalues and eigenvectors? Suppose that a basis has been chosen for the n-dimensional vector space X. Then the matrix representation for Eq. (6.46) can be written

$$\mathbf{Az} = \lambda\mathbf{z}, \qquad (6.47)$$

or

$$[\mathbf{A} - \lambda\mathbf{I}]\mathbf{z} = \mathbf{0}. \qquad (6.48)$$

This means that the columns of $[\mathbf{A} - \lambda\mathbf{I}]$ are dependent, and therefore the determinant of this matrix must be zero:

$$|[\mathbf{A} - \lambda\mathbf{I}]| = 0. \qquad (6.49)$$

This determinant is an nth-order polynomial. Therefore Eq. (6.49) always has n roots, some of which may be complex and some of which may be repeated.

As an example, let's revisit the rotation example. If we use the standard basis set, the matrix of the transformation is

$$\mathbf{A} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \qquad (6.50)$$

We can then write Eq. (6.49) as

$$\left|\begin{bmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{bmatrix}\right| = 0, \qquad (6.51)$$

or

$$\lambda^2 - 2\lambda\cos\theta + ((\cos\theta)^2 + (\sin\theta)^2) = \lambda^2 - 2\lambda\cos\theta + 1 = 0. \qquad (6.52)$$

The roots of this equation are

$$\lambda_1 = \cos\theta + j\sin\theta \qquad \lambda_2 = \cos\theta - j\sin\theta . \tag{6.53}$$

Therefore, as we predicted, this transformation has no real eigenvalues (if $\sin\theta \neq 0$). This means that when any real vector is transformed, it will point in a new direction.

Consider another matrix:

$$\mathbf{A} = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} . \tag{6.54}$$

To find the eigenvalues we must solve

$$\left| \begin{bmatrix} -1-\lambda & 1 \\ 0 & -2-\lambda \end{bmatrix} \right| = 0 , \tag{6.55}$$

or

$$\lambda^2 + 3\lambda + 2 = (\lambda + 1)(\lambda + 2) = 0 , \tag{6.56}$$

and the eigenvalues are

$$\lambda_1 = -1 \qquad \lambda_2 = -2 . \tag{6.57}$$

To find the eigenvectors we must solve Eq. (6.48), which in this example becomes

$$\begin{bmatrix} -1-\lambda & 1 \\ 0 & -2-\lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} . \tag{6.58}$$

We will solve this equation twice, once using λ_1 and once using λ_2. Beginning with λ_1 we have

$$\begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \mathbf{z}_1 = \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \tag{6.59}$$

or

$$z_{21} = 0 \text{, no constraint on } z_{11} . \tag{6.60}$$

Therefore the first eigenvector will be

$$\mathbf{z}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{6.61}$$

or any scalar multiple. For the second eigenvector we use λ_2:

$$\begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \mathbf{z}_2 = \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} z_{12} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \tag{6.62}$$

or

$$z_{22} = -z_{12} . \tag{6.63}$$

Therefore the second eigenvector will be

$$\mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \tag{6.64}$$

or any scalar multiple.

To verify our results we consider the following:

$$\mathbf{A}\mathbf{z}_1 = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} = (-1) \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \lambda_1 \mathbf{z}_1 , \tag{6.65}$$

$$\mathbf{A}\mathbf{z}_2 = \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \begin{bmatrix} -2 \\ 2 \end{bmatrix} = (-2) \begin{bmatrix} 1 \\ -1 \end{bmatrix} = \lambda_2 \mathbf{z}_2 . \tag{6.66}$$

To test your understanding of eigenvectors, use the Neural Network Design Demonstration Eigenvector Game (`nnd6eg`).

Diagonalization

Whenever we have n distinct eigenvalues we are guaranteed that we can find n independent eigenvectors [Brog91]. Therefore the eigenvectors make up a basis set for the vector space of the transformation. Let's find the matrix of the previous transformation (Eq. (6.54)) using the eigenvectors as the basis vectors. From Eq. (6.33) we have

$$\mathbf{A}' = [\mathbf{B}^{-1}\mathbf{A}\mathbf{B}] = \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} -1 & 1 \\ 0 & -2 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & -1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -2 \end{bmatrix} . \tag{6.67}$$

Note that this is a diagonal matrix, with the eigenvalues on the diagonal. This is not a coincidence. Whenever we have distinct eigenvalues we can diagonalize the matrix representation by using the eigenvectors as the ba-

Diagonalization

sis vectors. This *diagonalization* process is summarized in the following. Let

$$\mathbf{B} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_n \end{bmatrix}, \quad (6.68)$$

where $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ are the eigenvectors of a matrix $\mathbf{A}$. Then

$$[\mathbf{B}^{-1}\mathbf{A}\mathbf{B}] = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}, \quad (6.69)$$

where $\{\lambda_1, \lambda_2, \dots, \lambda_n\}$ are the eigenvalues of the matrix $\mathbf{A}$.

This result will be very helpful as we analyze the performance of several neural networks in later chapters.

Summary of Results

Transformations

A *transformation* consists of three parts:

1. a set of elements $X = \{\chi_i\}$, called the domain,
2. a set of elements $Y = \{y_i\}$, called the range, and
3. a rule relating each $\chi_i \in X$ to an element $y_i \in Y$.

Linear Transformations

A transformation $\mathcal{A}$ is *linear* if:

1. for all $\chi_1, \chi_2 \in X$, $\mathcal{A}(\chi_1 + \chi_2) = \mathcal{A}(\chi_1) + \mathcal{A}(\chi_2)$,
2. for all $\chi \in X$, $a \in R$, $\mathcal{A}(a\chi) = a\mathcal{A}(\chi)$.

Matrix Representations

Let $\{v_1, v_2, \ldots, v_n\}$ be a basis for vector space X, and let $\{u_1, u_2, \ldots, u_m\}$ be a basis for vector space Y. Let $\mathcal{A}$ be a linear transformation with domain X and range Y:

$$\mathcal{A}(\chi) = y.$$

The coefficients of the matrix representation are obtained from

$$\mathcal{A}(v_j) = \sum_{i=1}^{m} a_{ij} u_i .$$

Change of Basis

$$\mathbf{B}_t = \begin{bmatrix} \mathbf{t}_1 & \mathbf{t}_2 & \ldots & \mathbf{t}_n \end{bmatrix}$$

$$\mathbf{B}_w = \begin{bmatrix} \mathbf{w}_1 & \mathbf{w}_2 & \ldots & \mathbf{w}_m \end{bmatrix}$$

$$\mathbf{A}' = [\mathbf{B}_w^{-1} \mathbf{A} \mathbf{B}_t]$$

Eigenvalues and Eigenvectors

$$\mathbf{A}\mathbf{z} = \lambda \mathbf{z}$$

$$|[\mathbf{A} - \lambda \mathbf{I}]| = 0$$

Diagonalization

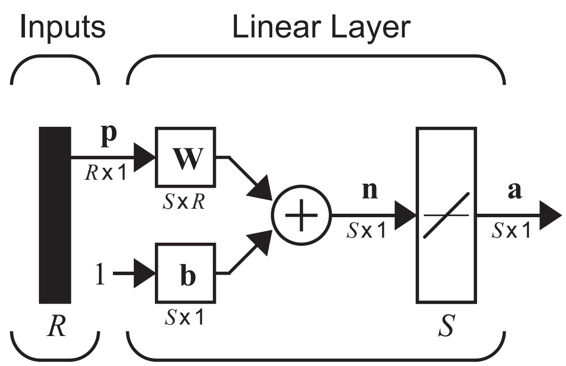
$$\mathbf{B} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 & \dots & \mathbf{z}_n \end{bmatrix},$$

where $\{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_n\}$ are the eigenvectors of a square matrix $\mathbf{A}$.

$$[\mathbf{B}^{-1}\mathbf{A}\mathbf{B}] = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Solved Problems

P6.1 Consider the single-layer network shown in Figure P6.1, which has a linear transfer function. Is the transformation from the input vector to the output vector a linear transformation?



$\mathbf{a} = \mathbf{purelin}(\mathbf{Wp} + \mathbf{b})$

Figure P6.1 Single-Neuron Perceptron

The network equation is

$$\mathbf{a} = \mathcal{A}(\mathbf{p}) = \mathbf{Wp} + \mathbf{b}.$$

In order for this transformation to be linear it must satisfy

1. $\mathcal{A}(\mathbf{p}_1 + \mathbf{p}_2) = \mathcal{A}(\mathbf{p}_1) + \mathcal{A}(\mathbf{p}_2)$,
2. $\mathcal{A}(a\mathbf{p}) = a\mathcal{A}(\mathbf{p})$.

Let's test condition 1 first.

$$\mathcal{A}(\mathbf{p}_1 + \mathbf{p}_2) = \mathbf{W}(\mathbf{p}_1 + \mathbf{p}_2) + \mathbf{b} = \mathbf{Wp}_1 + \mathbf{Wp}_2 + \mathbf{b}.$$

Compare this with

$$\mathcal{A}(\mathbf{p}_1) + \mathcal{A}(\mathbf{p}_2) = \mathbf{Wp}_1 + \mathbf{b} + \mathbf{Wp}_2 + \mathbf{b} = \mathbf{Wp}_1 + \mathbf{Wp}_2 + 2\mathbf{b}.$$

Clearly these two expressions will be equal only if $\mathbf{b} = \mathbf{0}$. Therefore this network performs a nonlinear transformation, even though it has a linear transfer function. This particular type of nonlinearity is called an affine transformation.

P6.2 We discussed projections in Chapter 5. Is a projection a linear transformation?

The projection of a vector χ onto a vector v is computed as

$$y = \mathcal{A}(\chi) = \frac{(\chi, v)}{(v, v)} v,$$

where (χ, v) is the inner product of χ with v.

We need to check to see if this transformation satisfies the two conditions for linearity. Let's start with condition 1:

$$\mathcal{A}(\chi_1 + \chi_2) = \frac{(\chi_1 + \chi_2, v)}{(v, v)} v = \frac{(\chi_1, v) + (\chi_2, v)}{(v, v)} v = \frac{(\chi_1, v)}{(v, v)} v + \frac{(\chi_2, v)}{(v, v)} v$$

$$= \mathcal{A}(\chi_1) + \mathcal{A}(\chi_2).$$

(Here we used linearity properties of inner products.) Checking condition 2:

$$\mathcal{A}(a\chi) = \frac{(a\chi, v)}{(v, v)} v = \frac{a(\chi, v)}{(v, v)} v = a\mathcal{A}(\chi).$$

Therefore projection is a linear operation.

P6.3 Consider the transformation $\mathcal{A}$ created by reflecting a vector χ in $\Re^2$ about the line $x_1 + x_2 = 0$, as shown in Figure P6.2. Find the matrix of this transformation relative to the standard basis in $\Re^2$.

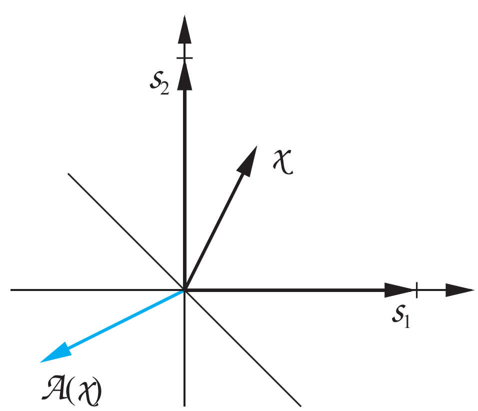
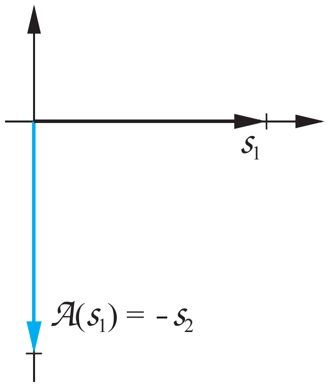


Figure P6.2 Reflection Transformation

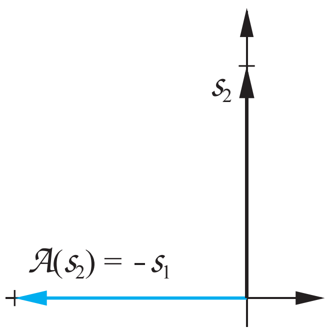
The key to finding the matrix of a transformation is given in Eq. (6.6):

$$\mathcal{A}(v_j) = \sum_{i=1}^{m} a_{ij} u_i .$$



We need to transform each basis vector of the domain and then expand the result in terms of the basis vectors for the range. Each time we do the expansion we get one column of the matrix representation. In this case the basis set for both the domain and the range is $\{s_1, s_2\}$. So let's transform s_1 first. If we reflect s_1 about the line $x_1 + x_2 = 0$, we find

$$\mathcal{A}(s_1) = -s_2 = \sum_{i=1}^{2} a_{i1} s_i = a_{11} s_1 + a_{21} s_2 = 0 s_1 + (-1) s_2$$



(as shown in the top left figure), which gives us the first column of the matrix. Next we transform s_2:

$$\mathcal{A}(s_2) = -s_1 = \sum_{i=1}^{2} a_{i2} s_i = a_{12} s_1 + a_{22} s_2 = (-1) s_1 + 0 s_2$$

(as shown in the second figure on the left), which gives us the second column of the matrix. The final result is

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

Let's test our result by transforming the vector $\mathbf{x} = \begin{bmatrix} 1 & 1 \end{bmatrix}^T$:

$$\mathbf{Ax} = \begin{bmatrix} 0 & -1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}.$$

This is indeed the reflection of $\mathbf{x}$ about the line $x_1 + x_2 = 0$, as we can see in Figure P6.3.

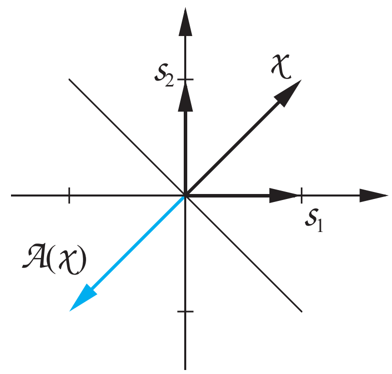


Figure P6.3 Test of Reflection Operation

(Can you guess the eigenvalues and eigenvectors of this transformation? Use the *Neural Network Design Demonstration Linear Transformations* (`nnd6lt`) to investigate this graphically. Compute the eigenvalues and eigenvectors, using the MATLAB function **`eig`**, and check your guess.)

P6.4 Consider the space of complex numbers. Let this be the vector space X, and let the basis for X be $\{1+j, 1-j\}$. Let $\mathcal{A}: X \rightarrow X$ be the conjugation operator (i.e., $\mathcal{A}(\chi) = \chi^*$).

i. Find the matrix of the transformation $\mathcal{A}$ relative to the basis set given above.

ii. Find the eigenvalues and eigenvectors of the transformation.

iii. Find the matrix representation for $\mathcal{A}$ relative to the eigenvectors as the basis vectors.

i. To find the matrix of the transformation, transform each of the basis vectors (by finding their conjugate):

$$\mathcal{A}(v_1) = \mathcal{A}(1+j) = 1-j = v_2 = a_{11} v_1 + a_{21} v_2 = 0 v_1 + 1 v_2 \,,$$

$$\mathcal{A}(v_2) = \mathcal{A}(1-j) = 1+j = v_1 = a_{12} v_1 + a_{22} v_2 = 1 v_1 + 0 v_2 \,.$$

This gives us the matrix representation

$$\mathbf{A} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} .$$

ii. To find the eigenvalues, we need to use Eq. (6.49):

$$|[\mathbf{A} - \lambda\mathbf{I}]| = \left|\begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}\right| = \lambda^2 - 1 = (\lambda - 1)(\lambda + 1) = 0 .$$

So the eigenvalues are: $\lambda_1 = 1$, $\lambda_2 = -1$. To find the eigenvectors, use Eq. (6.48):

$$[\mathbf{A} - \lambda\mathbf{I}]\mathbf{z} = \begin{bmatrix} -\lambda & 1 \\ 1 & -\lambda \end{bmatrix}\mathbf{z} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} .$$

For $\lambda = \lambda_1 = 1$ this gives us

$$\begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\mathbf{z}_1 = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} ,$$

or

$$z_{11} = z_{21} .$$

Therefore the first eigenvector will be

$$\mathbf{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} ,$$

or any scalar multiple. For the second eigenvector we use $\lambda = \lambda_2 = -1$:

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\mathbf{z}_1 = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} z_{12} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} ,$$

or

$$z_{12} = -z_{22} .$$

Therefore the second eigenvector is

$$\mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix} ,$$

or any scalar multiple.

Note that while these eigenvectors can be represented as columns of numbers, in reality they are complex numbers. For example:

$$z_1 = 1\,v_1 + 1\,v_2 = (1 + j) + (1 - j) = 2 ,$$

$$z_2 = 1v_1 + (-1)v_2 = (1+j) - (1-j) = 2j\,.$$

Checking that these are indeed eigenvectors:

$$\mathcal{A}(z_1) = (2)^* = 2 = \lambda_1 z_1\,,$$

$$\mathcal{A}(z_2) = (2j)^* = -2j = \lambda_2 z_2\,.$$

iii. To perform a change of basis we need to use Eq. (6.33):

$$\mathbf{A}' = [\mathbf{B}_w^{-1}\mathbf{A}\mathbf{B}_t] = [\mathbf{B}^{-1}\mathbf{A}\mathbf{B}]\,,$$

where

$$\mathbf{B} = \begin{bmatrix}\mathbf{z}_1 & \mathbf{z}_2\end{bmatrix} = \begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix}.$$

(We are using the same basis set for the range and the domain.) Therefore we have

$$\mathbf{A}' = \begin{bmatrix}0.5 & 0.5\\ 0.5 & -0.5\end{bmatrix}\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 1 & -1\end{bmatrix} = \begin{bmatrix}1 & 0\\ 0 & -1\end{bmatrix} = \begin{bmatrix}\lambda_1 & 0\\ 0 & \lambda_2\end{bmatrix}.$$

As expected from Eq. (6.69), we have diagonalized the matrix representation.

P6.5 Diagonalize the following matrix:

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

The first step is to find the eigenvalues:

$$|[\mathbf{A} - \lambda\mathbf{I}]| = \left|\begin{bmatrix}2-\lambda & -2\\ -1 & 3-\lambda\end{bmatrix}\right| = \lambda^2 - 5\lambda + 4 = (\lambda-1)(\lambda-4) = 0\,,$$

so the eigenvalues are $\lambda_1 = 1$, $\lambda_2 = 4$. To find the eigenvectors,

$$[\mathbf{A} - \lambda\mathbf{I}]\mathbf{z} = \begin{bmatrix}2-\lambda & -2\\ -1 & 3-\lambda\end{bmatrix}\mathbf{z} = \begin{bmatrix}0\\ 0\end{bmatrix}.$$

For $\lambda = \lambda_1 = 1$

$$\begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \mathbf{z}_1 = \begin{bmatrix} 1 & -2 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} z_{11} \\ z_{21} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

or

$$z_{11} = 2z_{21}.$$

Therefore the first eigenvector will be

$$\mathbf{z}_1 = \begin{bmatrix} 2 \\ 1 \end{bmatrix},$$

or any scalar multiple.

For $\lambda = \lambda_2 = 4$

$$\begin{bmatrix} -2 & -2 \\ -1 & -1 \end{bmatrix} \mathbf{z}_1 = \begin{bmatrix} -2 & -2 \\ -1 & -1 \end{bmatrix} \begin{bmatrix} z_{12} \\ z_{22} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

or

$$z_{12} = -z_{22}.$$

Therefore the second eigenvector will be

$$\mathbf{z}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix},$$

or any scalar multiple.

To diagonalize the matrix we use Eq. (6.69):

$$\mathbf{A}' = [\mathbf{B}^{-1}\mathbf{A}\mathbf{B}],$$

where

$$\mathbf{B} = \begin{bmatrix} \mathbf{z}_1 & \mathbf{z}_2 \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix}.$$

Therefore we have

$$\mathbf{A}' = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ \frac{1}{3} & -\frac{2}{3} \end{bmatrix} \begin{bmatrix} 2 & -2 \\ -1 & 3 \end{bmatrix} \begin{bmatrix} 2 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 4 \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix}.$$

P6.6 Consider a transformation $\mathcal{A}: R^3 \to R^2$ whose matrix representation relative to the standard basis sets is

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

Find the matrix for this transformation relative to the basis sets:

$$T = \left\{ \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}, \begin{bmatrix} 0 \\ -1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \\ 3 \end{bmatrix} \right\} \qquad W = \left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 0 \\ -2 \end{bmatrix} \right\}.$$

The first step is to form the matrices

$$\mathbf{B}_t = \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & -2 \\ 1 & 0 & 3 \end{bmatrix} \qquad \mathbf{B}_w = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}.$$

Now we use Eq. (6.33) to form the new matrix representation:

$$\mathbf{A}' = [\mathbf{B}_w^{-1} \mathbf{A} \mathbf{B}_t],$$

$$\mathbf{A}' = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 3 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & -1 & -2 \\ 1 & 0 & 3 \end{bmatrix} = \begin{bmatrix} 6 & 1 & 2 \\ -\frac{1}{2} & 0 & -\frac{3}{2} \end{bmatrix}.$$

Therefore this is the matrix of the transformation with respect to the basis sets T and W.

P6.7 Consider a transformation $\mathcal{A}: \Re^2 \to \Re^2$. One basis set for $\Re^2$ is given as $V = \{v_1, v_2\}$.

i. Find the matrix of the transformation $\mathcal{A}$ relative to the basis set V if it is given that

$$\mathcal{A}(v_1) = v_1 + 2v_2,$$

$$\mathcal{A}(v_2) = v_1 + v_2.$$

ii. Consider a new basis set $W = \{w_1, w_2\}$. Find the matrix of the transformation $\mathcal{A}$ relative to the basis set W if it is given that

$$w_1 = v_1 + v_2,$$

$$w_2 = v_1 - v_2.$$

i. Each of the two equations gives us one column of the matrix, as defined in Eq. (6.6). Therefore the matrix is

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

ii. We can represent the W basis vectors as columns of numbers in terms of the V basis vectors:

$$\mathbf{w}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \mathbf{w}_2 = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

We can now form the basis matrix that we need to perform the similarity transform:

$$\mathbf{B}_w = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

The new matrix representation can then be obtained from Eq. (6.33):

$$\mathbf{A}' = [\mathbf{B}_w^{-1}\mathbf{A}\mathbf{B}_w],$$

$$\mathbf{A}' = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} \frac{5}{2} & \frac{1}{2} \\ -\frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

P6.8 Consider the vector space P^2 of all polynomials of degree less than or equal to 2. One basis for this vector space is $V = \{1, t, t^2\}$. Consider the differentiation transformation $\mathcal{D}$.

i. Find the matrix of this transformation relative to the basis set V.

ii. Find the eigenvalues and eigenvectors of the transformation.

i. The first step is to transform each of the basis vectors:

$$\mathcal{D}(1) = 0 = (0)1 + (0)t + (0)t^2 ,$$

$$\mathcal{D}(t) = 1 = (1)1 + (0)t + (0)t^2 ,$$

$$\mathcal{D}(t^2) = 2t = (0)1 + (2)t + (0)t^2 .$$

The matrix of the transformation is then given by

$$\mathbf{D} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} .$$

ii. To find the eigenvalues we must solve

$$|[\mathbf{D} - \lambda\mathbf{I}]| = \left| \begin{bmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 2 \\ 0 & 0 & -\lambda \end{bmatrix} \right| = -\lambda^3 = 0 .$$

Therefore all three eigenvalues are zero. To find the eigenvectors we need to solve

$$[\mathbf{D} - \lambda\mathbf{I}]\mathbf{z} = \begin{bmatrix} -\lambda & 1 & 0 \\ 0 & -\lambda & 2 \\ 0 & 0 & -\lambda \end{bmatrix} \mathbf{z} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} .$$

For $\lambda = 0$ we have

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z_1 \\ z_2 \\ z_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} .$$

This means that

$$z_2 = z_3 = 0 .$$

Therefore we have a single eigenvector:

$$\mathbf{z} = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} .$$

Therefore the only polynomial whose derivative is a scaled version of itself is a constant (a zeroth-order polynomial).

P6.9 Consider a transformation $\mathcal{A}: R^2 \rightarrow R^2$. Two examples of transformed vectors are given in Figure P6.4. Find the matrix representation of this transformation relative to the standard basis set.

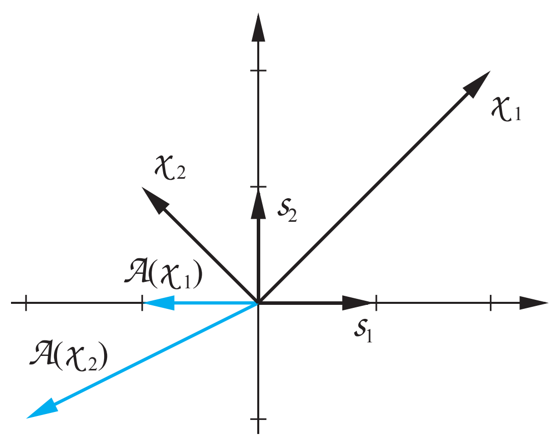


Figure P6.4 Transformation for Problem P6.9

For this problem we do not know how the basis vectors are transformed, so we cannot use Eq. (6.6) to find the matrix representation. However, we do know how two vectors are transformed, and we do know how those vectors can be represented in terms of the standard basis set. From Figure P6.4 we can write the following equations:

$$\mathbf{A}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix} , \mathbf{A}\begin{bmatrix} -1 \\ 1 \end{bmatrix} = \begin{bmatrix} -2 \\ -1 \end{bmatrix} .$$

We then put these two equations together to form

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

So that

$$\mathbf{A} = \begin{bmatrix} -1 & -2 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} 2 & -1 \\ 2 & 1 \end{bmatrix}^{-1} = \begin{bmatrix} -1 & -2 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & \frac{1}{4} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix} = \begin{bmatrix} \frac{3}{4} & -\frac{5}{4} \\ \frac{1}{2} & -\frac{1}{2} \end{bmatrix}.$$

This is the matrix representation of the transformation with respect to the standard basis set.

This procedure is used in the *Neural Network Design Demonstration Linear Transformations* (`nnd6lt`).

Epilogue

In this chapter we have reviewed those properties of linear transformations and matrices that are most important to our study of neural networks. The concepts of eigenvalues, eigenvectors, change of basis (similarity transformation) and diagonalization will be used again and again throughout the remainder of this text. Without this linear algebra background our study of neural networks could only be superficial.

In the next chapter we will use linear algebra to analyze the operation of one of the first neural network training algorithms — the Hebb rule.

Further Reading

[Brog91] W. L. Brogan, *Modern Control Theory*, 3rd Ed., Englewood Cliffs, NJ: Prentice-Hall, 1991.

This is a well-written book on the subject of linear systems. The first half of the book is devoted to linear algebra. It also has good sections on the solution of linear differential equations and the stability of linear and nonlinear systems. It has many worked problems.

[Stra76] G. Strang, *Linear Algebra and Its Applications*, New York: Academic Press, 1980.

Strang has written a good basic text on linear algebra. Many applications of linear algebra are integrated into the text.

Exercises

E6.1 Is the operation of transposing a matrix a linear transformation?

E6.2 Consider again the neural network shown in Figure P6.1. Show that if the bias vector $\mathbf{b}$ is equal to zero then the network performs a linear operation.

E6.3 Consider the linear transformation illustrated in Figure E6.1.

i. Find the matrix representation of this transformation relative to the standard basis set.

ii. Find the matrix of this transformation relative to the basis set $\{v_1, v_2\}$.

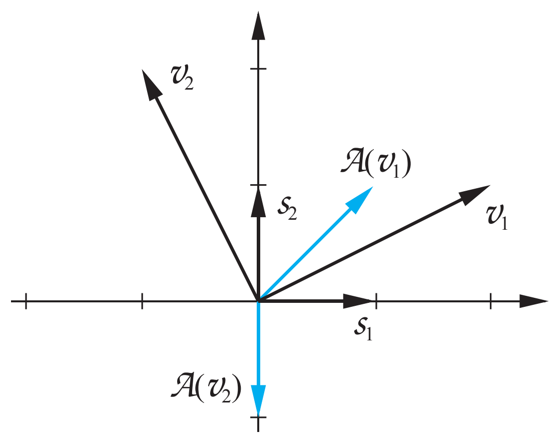


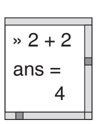
Figure E6.1 Example Transformation for Exercise E6.3

E6.4 Consider the space of complex numbers. Let this be the vector space X, and let the basis for X be $\{1+j, 1-j\}$. Let $\mathcal{A}: X \rightarrow X$ be the operation of multiplication by $(1+j)$ (i.e., $\mathcal{A}(\chi) = (1+j)\chi$).

i. Find the matrix of the transformation $\mathcal{A}$ relative to the basis set given above.

ii. Find the eigenvalues and eigenvectors of the transformation.

iii. Find the matrix representation for $\mathcal{A}$ relative to the eigenvectors as the basis vectors.



iv. Check your answers to parts (ii) and (iii) using MATLAB.

E6.5 Consider a transformation $\mathcal{A}:P^2 \to P^3$, from the space of second-order polynomials to the space of third-order polynomials, which is defined by the following:

$$\chi = a_0 + a_1 t + a_2 t^2 ,$$

$$\mathcal{A}(\chi) = a_0(t+1) + a_1(t+1)^2 + a_2(t+1)^3 .$$

Find the matrix representation of this transformation relative to the basis sets $V^2 = \{1, t, t^2\}$, $V^3 = \{1, t, t^2, t^3\}$.

E6.6 Consider the vector space of polynomials of degree two or less. These polynomials have the form $f(t) = a_0 + a_1 t + a_2 t^2$. Now consider the transformation in which the variable t is replaced by $t+1$. (for example, $t^2 + 2t + 3 \Rightarrow (t+1)^2 + 2(t+1) + 3 = t^2 + 4t + 6$)

i. Find the matrix of this transformation with respect to the basis set $\{1, t-1, t^2\}$.

ii. Find the eigenvalues and eigenvectors of the transformation. Show the eigenvectors as columns of numbers and as functions of time (polynomials).

E6.7 Consider the space of functions of the form $\alpha \sin(t + \phi)$. One basis set for this space is $V = \{\sin t, \cos t\}$. Consider the differentiation transformation $\mathcal{D}$.

i. Find the matrix of the transformation $\mathcal{D}$ relative to the basis set V.

ii. Find the eigenvalues and eigenvectors of the transformation. Show the eigenvectors as columns of numbers and as functions of t.

iii. Find the matrix of the transformation relative to the eigenvectors as basis vectors.

E6.8 Consider the vector space of functions of the form $\alpha + \beta e^{2t}$. One basis set for this vector space is $V = \{1 + e^{2t}, 1 - e^{2t}\}$. Consider the differentiation transformation $\mathcal{D}$.

i. Find the matrix of the transformation $\mathcal{D}$ relative to the basis set V, using Eq. (6.6).

ii. Verify the operation of the matrix on the function $2e^{2t}$.

iii. Find the eigenvalues and eigenvectors of the transformation. Show the eigenvectors as columns of numbers (with respect to the basis set V) and as functions of t.

iv. Find the matrix of the transformation relative to the eigenvectors as basis vectors.

E6.9 Consider the set of all 2x2 matrices. This set is a vector space, which we will call X (yes, matrices can be vectors). If $\mathbf{M}$ is an element of this vector space, define the transformation $\mathcal{A}: X \to X$, such that $\mathcal{A}(\mathbf{M}) = \mathbf{M} + \mathbf{M}^T$. Consider the following basis set for the vector space X.

$$v_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix},\ v_2 = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix},\ v_3 = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix},\ v_4 = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}$$

i. Find the matrix representation of the transformation $\mathcal{A}$ relative to the basis set $\{v_1, v_2, v_3, v_4\}$ (for both domain and range) (using Eq. (6.6)).

ii. Verify the operation of the matrix representation from part i. on the element of X given below. (Verify that the matrix multiplication produces the same result as the transformation.)

$$\begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}$$

iii. Find the eigenvalues and eigenvectors of the transformation. You do not need to use the matrix representation that you found in part i. You can find the eigenvalues and eigenvectors directly from the definition of the transformation. Your eigenvectors should be 2x2 matrices (elements of the vector space X). This does not require much computation. Use the definition of eigenvector in Eq. (6.46).

E6.10 Consider a transformation $\mathcal{A}: P^1 \to P^2$, from the space of first degree polynomials into the space of second degree polynomials. The transformation is defined as follows

$$\mathcal{A}(a + bt) = at + \frac{b}{2}t^2$$

(e.g., $\mathcal{A}(2 + 6t) = 2t + 3t^2$). One basis set for P^1 is $U = \{1, t\}$. One basis for P^2 is $V = \{1, t, t^2\}$.

i. Find the matrix representation of the transformation A relative to the basis sets U and V, using Eq. (6.6).

ii. Verify the operation of the matrix on the polynomial $6 + 8t$. (Verify that the matrix multiplication produces the same result as the transformation.)

iii. Using a similarity transform, find the matrix of the transformation with respect to the basis sets $S = \{1 + t, 1 - t\}$ and V.

E6.11 Let $\mathcal{D}$ be the differentiation operator ($\mathcal{D}(f) = df/dt$), and use the basis set

$$\{u_1, u_2\} = \{e^{5t}, te^{5t}\}$$

for both the domain and the range of the transformation $\mathcal{D}$.

i. Show that the transformation $\mathcal{D}$ is linear.,

ii. Find the matrix of this transformation relative to the basis shown above.

iii. Find the eigenvalues and eigenvectors of the transformation $\mathcal{D}$.

E6.12 A certain linear transformation has the following eigenvalues and eigenvectors (represented in terms of the standard basis set):

$$\left\{\mathbf{z}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \lambda_1 = 1\right\}, \left\{\mathbf{z}_2 = \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \lambda_2 = 2\right\}$$

i. Find the matrix representation of the transformation, relative to the standard basis set.

ii. Find the matrix representation of the transformation relative to the eigenvectors as the basis vectors.

E6.13 Consider a transformation $\mathcal{A}: \Re^2 \rightarrow \Re^2$. In the figure below, we show a set of basis vectors $V = \{v_1, v_2\}$ and the transformed basis vectors.

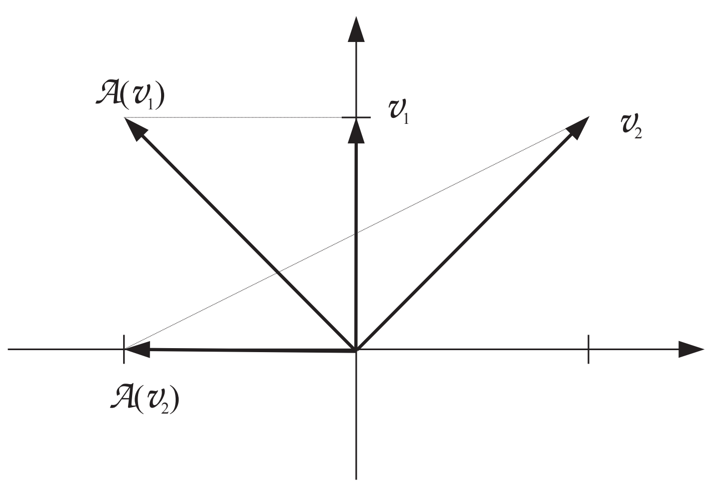


Figure E6.2 Definition of Transformation for Exercise E6.13

i. Find the matrix representation of this transformation with respect to the basis vectors $V = \{v_1, v_2\}$.

ii. Find the matrix representation of this transformation with respect to the standard basis vectors.

iii. Find the eigenvalues and eigenvectors of this transformation. Sketch the eigenvectors and their transformations.

iv. Find the matrix representation of this transformation with respect to the eigenvectors as the basis vectors.

E6.14 Consider the vector spaces P^2 and P^3 of second-order and third-order polynomials. Find the matrix representation of the integration transformation $I{:}P^2 \to P^3$, relative to the basis sets $V^2 = \{1, t, t^2\}$, $V^3 = \{1, t, t^2, t^3\}$.

E6.15 A certain linear transformation $\mathcal{A}{:}\Re^2 \to \Re^2$ has a matrix representation relative to the standard basis set of

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

Find the matrix representation of this transformation relative to the new basis set:

$$V = \left\{ \begin{bmatrix} 1 \\ 3 \end{bmatrix}, \begin{bmatrix} 2 \\ 5 \end{bmatrix} \right\}.$$

E6.16 We know that a certain linear transformation $\mathcal{A}{:}R^2 \to R^2$ has eigenvalues and eigenvectors given by

$$\lambda_1 = 1 \qquad \mathbf{z}_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad \lambda_2 = 2 \qquad \mathbf{z}_2 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}.$$

(The eigenvectors are represented relative to the standard basis set.)

i. Find the matrix representation of the transformation $\mathcal{A}$ relative to the standard basis set.

ii. Find the matrix representation relative to the new basis

$$V = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} -1 \\ 1 \end{bmatrix} \right\}.$$

E6.17 Consider the transformation $\mathcal{A}$ created by projecting a vector χ onto the line shown in Figure E6.3. An example of the transformation is shown in the figure.

i. Using Eq. (6.6), find the matrix representation of this transformation relative to the standard basis set $\{s_1, s_2\}$.

ii. Using your answer to part i, find the matrix representation of this transformation relative to the basis set $\{v_1, v_2\}$ shown in Figure E6.3.

iii. What are the eigenvalues and eigenvectors of this transformation? Sketch the eigenvectors and their transformations.

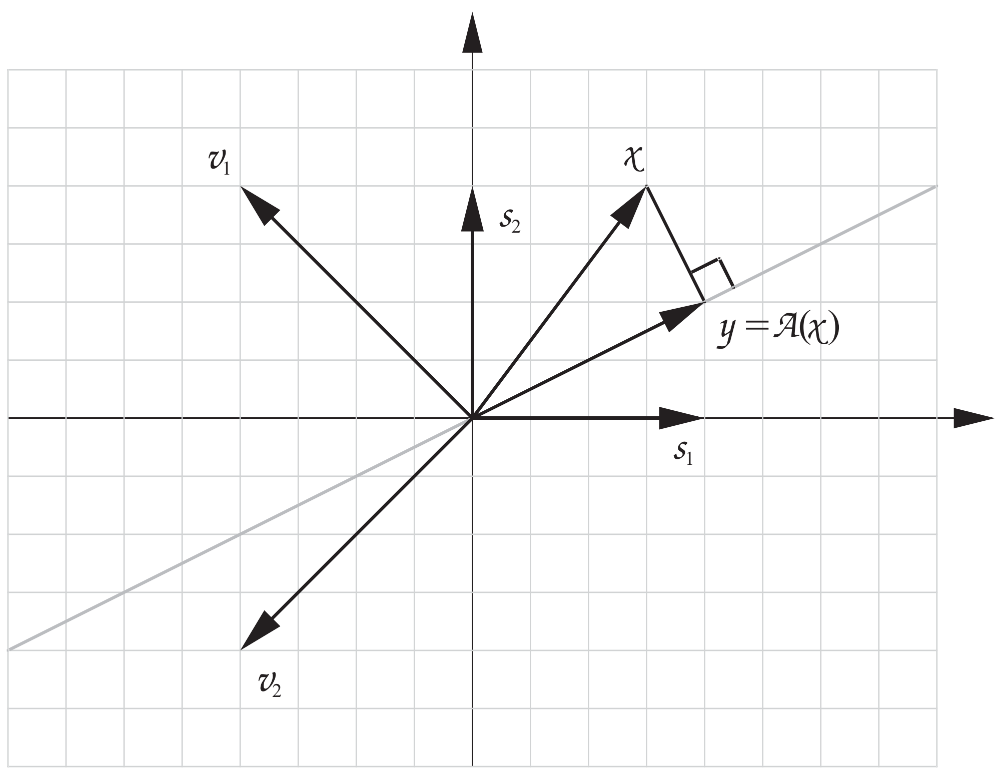


Figure E6.3 Definition of Transformation for Exercise E6.17

E6.18 Consider the following basis set for $\Re^2$:

$$V = \{\mathbf{v}_1, \mathbf{v}_2\} = \left\{ \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 1 \\ -2 \end{bmatrix} \right\}.$$

(The basis vectors are represented relative to the standard basis set.)

i. Find the reciprocal basis vectors for this basis set.

ii. Consider a transformation $\mathcal{A}: \Re^2 \to \Re^2$. The matrix representation for $\mathcal{A}$ relative to the standard basis in $\Re^2$ is

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

Find the expansion of $\mathbf{A}\mathbf{v}_1$ in terms of the basis set V. (Use the reciprocal basis vectors.)

iii. Find the expansion of $\mathbf{A}\mathbf{v}_2$ in terms of the basis set V.

iv. Find the matrix representation for $\mathcal{A}$ relative to the basis V. (This step should require no further computation.)