## Contents

### INTUITION

<table>
<thead>
<tr>
<th>CHAPTER 1</th>
<th>Functions</th>
<th>1</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1 Introduction</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>1.2 Partially defined functions</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>1.3 Vectors and Coordinates</td>
<td>7</td>
<td></td>
</tr>
<tr>
<td>1.4 Visualizing transformations of the plane</td>
<td>9</td>
<td></td>
</tr>
<tr>
<td>1.5 Vector fields</td>
<td>10</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 2</th>
<th>Coordinates</th>
<th>15</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.1 A generalization</td>
<td>15</td>
<td></td>
</tr>
<tr>
<td>2.2 Tayloring coordinates to the function</td>
<td>16</td>
<td></td>
</tr>
<tr>
<td>2.3 Changing coordinates</td>
<td>19</td>
<td></td>
</tr>
<tr>
<td>2.4 Running with the algebra</td>
<td>21</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 3</th>
<th>Matrices</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.1 The leading role</td>
<td>26</td>
<td></td>
</tr>
<tr>
<td>3.2 The matrix of a transformation</td>
<td>28</td>
<td></td>
</tr>
<tr>
<td>3.3 Matrix algebra</td>
<td>30</td>
<td></td>
</tr>
<tr>
<td>3.4 Some matrix forms</td>
<td>32</td>
<td></td>
</tr>
<tr>
<td>3.5 Rank and nullity</td>
<td>34</td>
<td></td>
</tr>
<tr>
<td>3.6 Invariants</td>
<td>36</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 4</th>
<th>Diagonalization</th>
<th>43</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1 An algorithm</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>4.2 The fundamental theorem of algebra</td>
<td>46</td>
<td></td>
</tr>
<tr>
<td>4.3 Eigenspaces</td>
<td>49</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 5</th>
<th>Projections</th>
<th>53</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.1 Plane projections</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>5.2 Space projections</td>
<td>57</td>
<td></td>
</tr>
<tr>
<td>5.3 An algebraic characterization of projections</td>
<td>61</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CHAPTER 6</th>
<th>Reflections et cetera</th>
<th>64</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2 Plane reflections</td>
<td>64</td>
<td></td>
</tr>
<tr>
<td>6.3 Space reflections</td>
<td>65</td>
<td></td>
</tr>
</tbody>
</table>
CHAPTER 12  Similarity

12.1 Matrix of a transformation  120
12.2 Coordinates and isomorphisms  121
12.3 Similarity  121
12.4 Upper triangular form  122
12.5 Two variants on upper triangular form  122

SOLUTIONS

CHAPTER 13  Bases

13.1 Vector Spaces  127

CHAPTER 14  Similarity

APPENDIX

APPENDIX 1  Complex Numbers

14.1 Endowing \( \mathbb{R}^2 \) with more structure  135
14.2 The linearity of multiplication  135
14.3 The traditional notation  135
14.4 Representations of fields  135
14.5 Factoring complex polynomials  136
14.6 Operations on complex numbers  136
14.7 Extending the domains of real functions  137

INDEX
Intuition
Chapter 1

Functions

1.1 Introduction

We assume the reader is familiar with numerical functions, like those encountered in a pre-calculus class. There we meet functions defined by formulae, such as $x^2$, $e^x$, and $\sin x$. Customarily we denote a function with a letter, like $f$, and this letter will be repeatedly recycled to denote different functions at different times. If we need to talk of two functions at the same time, two letters will be introduced to represent these functions; $f$ and $g$ being good candidates.

When we write the expression $f(x) = x^2$, it is understood that $x$ denotes an input of the function, $f(x)$ denotes the resulting output, while $f$ denotes the function itself. A function is really three things; two sets and one rule. The two sets are called the domain and codomain of the function, and the rule (or action) is frequently given by a formula, such as $f(x) = x^2 - \sin x$. In that formula, the symbol $x$ denotes an element of the domain, while $x^2 - \sin x$ finds itself in the codomain. We write

$$f : X \to Y$$

to concisely label a function $f$, its domain $X$ and codomain $Y$. In everything described above, it is the position of the symbols that is crucial. Writing $Y : f \to X$ communicates that $Y$ is a function having $f$ as its domain and $X$ as the codomain. In practice, authors attempt to maintain consistency in their choice of symbols, so you will almost always see $f$ denoting the function, not the domain of a function.

The domain of a function $f$ constitutes the set of inputs of $f$, and the set of actual outputs, i.e. the set of element of the form $f(x)$, is called the range of $f$. Beware that the range is not the same as the codomain. It is always true that the range is a subset of the codomain, but the range need not equal the codomain. With $f : X \to Y$, writing $f(x)$ implies that $x \in X$ and $f(x) \in Y$, and the range of $f$ is exactly

$$\{f(x) : x \in X\}.$$ 

This is always a subset of $Y$, but it might not equal $Y$. For example, if $f : R \to R$ is a constant function, then both the domain and codomain is the set of all real numbers $R$, but the range of $f$ has only one element.

Two functions $f$ and $g$ are equal exactly when three things happen: they have the same domain, the same codomain, and $f(x) = g(x)$ for each $x$ in the common domain.

**Example 1**

Suppose the rules of $f$ and $g$ are given by

$$f(x) = g(x) = \sqrt{x}$$

for all positive real numbers $x$. If the codomain of $f$ is $R$ and the codomain of $g$ is $\{x : x > 0\}$, then $f$ and $g$ are not equal according to our definition above. (The second condition for equality of functions fails, since the codomain assigned to $g$ is different than the one assigned to $f$.)
Chapter 1 • Functions

The previous example illustrates the central point we are making: two functions with the same rule can be different. The domain of both \( f \) and \( g \) is given implicitly when we say “for all positive real numbers \( x \)”. Alternatively, we could explicitly indicate both the domain and codomain by writing \( f : \{x : x > 0\} \rightarrow \mathbb{R} \).

Example 2

Assume \( s : N \rightarrow \mathbb{R} \) is given by \( s_n = n^2 \), and \( f : \mathbb{R} \rightarrow \mathbb{R} \) is given by \( f(x) = x^2 \). These functions are not equal because they have different domains.

The two functions in this example share a common action, that of squaring the input. When a function has a discrete domain, we sometimes write \( s_n \) instead of \( s(n) \), and when the domain is \( N \) such a function is called a sequence.

Example 3

If \( f : \mathbb{R} \rightarrow \mathbb{R}^2 \) and \( g : \mathbb{R} \rightarrow \{(x, x) : x \in \mathbb{R}\} \) are both defined by taking \( x \) to \((x, x)\), then the actions are identical, but the codomains are not equal so \( f \neq g \).

Here we see an example where the codomain of a function is different than the range of the function. Writing \( f : \mathbb{R} \rightarrow \mathbb{R}^2 \) explicitly declares that the codomain of \( f \) is the set \( \mathbb{R}^2 \), while the range of \( f \) (the actual outputs of \( f \)) is the set of all pairs of the form \((x, x)\), which looks like the line \( y = x \) inside \( \mathbb{R}^2 \). Thus \((1, 2)\) is not in the range but it is in the codomain, while the pair \((4, 4)\) is in both the range and the codomain of \( f \). On the other hand, the function \( g \) has the property that its range is equal to its codomain. When the range and codomain coincide for a function, we say that function is onto, so \( g \) is onto while \( f \) is not.

Example 4

Give an example of two functions \( f \) and \( g \) with common domain \( C \), and with identical rules, but \( f \neq g \).

The solution requires us to think of a rule that makes sense for all complex numbers. It makes sense to square complex numbers, so let that be the rule for both \( f \) and \( g \). The domain in common is then \( C \), the set of all complex numbers, and we need to invent codomains that are different (otherwise we would have \( f = g \)), and the codomains must contain the square of every complex number (because the codomain must contain the range). Those who know something about complex numbers realize that every complex number is a square, so the codomains for both \( f \) and \( g \) must contain all of \( C \), and they must be different. Let us set the codomain of \( f \) to be \( C \) itself, so now \( f \) is determined as a function (and it happens to be onto). To finish the solution, we need to find a set to serve as the codomain of \( g \), and it must contain \( C \), but not be equal to \( C \). This requires us to imagine something that is not a complex number, something such as \( \emptyset \), the empty set. We can then create a set \( Y = C \cup \{\emptyset\} \) to serve as the codomain of \( g \). This set \( Y \) consists of all the complex numbers, plus one more entity, the empty set. (It is perfectly legal for one set to be an element of some other set, although it is prohibited for a set to be an element of itself!)

Problem 1

Give examples of two functions \( r \) and \( s \) with common domain \( N \) and with identical rules, but \( r \neq s \).

Problem 2

Assume \( f : \{x : x > 0\} \rightarrow \{x : x > 0\} \) is defined by the rule \( f(x) = x^2 + 4x + 12 \). Indicate what the domain, codomain, and range of \( f \) is.

Problem 3

Give examples of \( f \) and \( g \) with common domain \( R \), common codomain \( \mathbb{R}^2 \), but whose ranges are different.

Problem 4

Give examples of \( f \) and \( g \) with common domain \( \mathbb{R}^2 \), common codomain \( \mathbb{R}^2 \), and common ranges, but \( f \neq g \).
Problem 5 Assume $f$ and $g$ have a common domain and identical actions. Show that $f$ and $g$ have the same range. Give an example to show that $f$ need not equal $g$.

Example 5 Let $f : C \to R^2$ be defined by $f(x + iy) = (x, y)$. The symbolism $f : C \to R^2$ tells us that the domain is the set of complex numbers, while the codomain is the Cartesian plane. The range of $f$ is the set of actual outputs, which is revealed by the formula $f(x + iy) = (x, y)$. Thus the range is

$$\{ f(x + iy) : x + iy \in C \} = \{ (x, y) : x, y \in R \} = R^2,$$

and $f$ is onto.

Problem 6 Assume $f : R^2 \to R^2$ is defined by $f(x, y) = (x, x)$. Give the range of $f$ and determine whether or not $f$ is onto.

Problem 7 Assume $f : R^2 \to R^2$ is defined by $f(x, y) = (x + y, x - y)$. Give the range of $f$ and determine whether or not $f$ is onto.

Consider the functions in problems 6 and 7; one of them has the property that two distinctly different inputs are taken to the same output. This can be written (as an equation) as $f(u) = f(v)$ with $u \neq v$. This is an example of a function that is not one-to-one. If there is no instance of two distinct inputs being mapped to the same output, we say the function is one-to-one. To test a function to see if it is one-to-one, set two arbitrary outputs equal and see if that equation forces the inputs equal. For example, we could write $f(u) = f(v)$ in problem 7 and proceed deductively, as follows. Realizing that both $u$ and $v$ are in the domain of $f$ we see they are elements of $R^2$, so $u = (x_1, y_1)$ and $v = (x_2, y_2)$, and the equality of outputs amounts to the equation $(x_1 + y_1, x_1 - y_1) = (x_2 + y_2, x_2 - y_2)$. Thus the first and second coordinates of the outputs are equal; i.e.

$$x_1 + y_1 = x_2 + y_2$$
$$x_1 - y_1 = x_2 - y_2.$$

If we add these two equations we are left with $2x_1 = 2x_2$, so the first coordinates of the inputs are equal. Subtracting the equations shows that the input’s second coordinates are also equal, and thus $u = v$. This constitutes a proof that the function given in problem 7 is one-to-one. The proof that the function in problem 6 is not one-to-one is simpler: one just gives an example of two vectors with $u \neq v$, but $f(u) = f(v)$. Such an example is called a counterexample, and a working counterexample for problem 6 is given by $u = (1, 2)$ and $v = (1, 3)$. Notice that there are infinitely many counterexamples. When giving a proof via a counterexample it is not polite to give all counterexamples. One counterexample is sufficient.

Sometimes we do not know whether we should try to prove a function is one-to-one, or whether we should give a counterexample, because we do not know beforehand whether the function is one-to-one or not. What is done in this case is experimentation. We try to prove it is one-to-one. If we get stuck, we look for counterexamples. If we can not find any, we again try to prove that it is one-to-one. This back and forth technique is much like solving a puzzle: if we try proving a function is one-to-one when it is in fact not one-to-one, our deduction should hit a snag, and a counterexample usually resides near that snag.

Problem 8 Assume $f : R^2 \to R^2$ is defined by $f(x, y) = (x + y, x + y)$. Give the range of $f$ and determine whether or not $f$ is onto. Show that $f$ is not one-to-one.
Example 6

Show that the function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by

\[
f(x, y) = (x + y, y)
\]

is one-to-one.

Our solution starts by writing \( f(x_1, y_1) = f(x_2, y_2) \), then we substitute to get

\[
(x_1 + y_1, y_1) = (x_2 + y_2, y_2).
\]

Equating the first and second coordinates gives us \( x_1 + y_1 = x_2 + y_2 \) and \( y_1 = y_2 \). Thus we see the second coordinates are equal, and substituting this fact into the equation \( x_1 + y_1 = x_2 + y_2 \) tells us the first coordinates are equal, enabling us to conclude \( (x_1, y_1) = (x_2, y_2) \).

Problem 9

Assume \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is defined by \( f(x, y) = (x, x + y) \). Show that \( f \) is one-to-one.

Problem 10

Assume \( f : \mathbb{C} \to \mathbb{C} \) is defined by \( f(x + iy) = x^2 + y^2 \). Show that \( f \) is not one-to-one.

Problem 11

Assume \( f : \mathbb{C} \to \mathbb{C} \) is defined by \( f(x + iy) = (x + iy)^2 \). Show that \( f \) is not one-to-one.

Problem 12

Assume \( f : \mathbb{C} \to \mathbb{C} \) is defined by \( f(w) = w^3 \). Show that \( f \) is not one-to-one.

It is important to realize that the notions of one-to-one and onto depend on the domain and codomain of a function, not just the rule that defines the function. When \( f : \mathbb{R} \to \mathbb{R} \) is defined by \( f(w) = w^3 \), then \( f \) is one-to-one, while the function in problem 12 is not.

A function \( f : X \to Y \) is invertible exactly when \( f \) is both one-to-one and onto. When this is the case \( f \) has an inverse function: i.e. there is a function \( g : Y \to X \) so that \( g(f(x)) = x \) and \( f(g(y)) = y \) for all \( x \in X \) and \( y \in Y \). You will find a list of invertible functions in Table 1. Notice that explicitly mentioning the domain and codomain of the function is essential to the definition of the function, and the question of whether the function is invertible can not be answered without this information. For example, the function \( f : \mathbb{R} \to \mathbb{R} \) defined by \( f(x) = x^2 \) is not invertible, while the function \( f : [0, \infty) \to [0, \infty) \) defined by \( f(x) = x^2 \) is invertible (it’s inverse is \( g(x) = \sqrt{x} \)).

<table>
<thead>
<tr>
<th>Function</th>
<th>Inverse</th>
</tr>
</thead>
<tbody>
<tr>
<td>( f : [0, \infty) \to [0, \infty) ) ( f(x) = x^2 )</td>
<td>( g : [0, \infty) \to [0, \infty) ) ( g(x) = \sqrt{x} )</td>
</tr>
<tr>
<td>( f : \mathbb{R} \to \mathbb{R} ) ( f(x) = x^3 )</td>
<td>( g : \mathbb{R} \to \mathbb{R} ) ( g(x) = x^{1/3} )</td>
</tr>
<tr>
<td>( f : [-1, 1] \to [-\frac{\pi}{2}, \frac{\pi}{2}] ) ( f(x) = \arcsin x )</td>
<td>( g : [-\frac{\pi}{2}, \frac{\pi}{2}] \to [-1, 1] ) ( g(x) = \sin x )</td>
</tr>
<tr>
<td>( f : \mathbb{R} \to (0, \infty) ) ( f(x) = 10^x )</td>
<td>( g : (0, \infty) \to \mathbb{R} ) ( g(x) = \log_{10} x )</td>
</tr>
<tr>
<td>( f : \mathbb{R} \to (0, \infty) ) ( f(x) = e^x )</td>
<td>( g : (0, \infty) \to \mathbb{R} ) ( g(x) = \ln x )</td>
</tr>
</tbody>
</table>

Table 1: Functions and their inverses.

On any set \( X \) there is a function \( \iota : X \to X \) that does nothing; if \( x \in X \) then \( \iota(x) = x \). This is called the identity function, and the word ‘identity’ is used because the numerical identity ‘1’ is to multiplication and other numbers exactly what the identity function ‘\( \iota \)’ is to composition and other functions. Composition is a way of combining two functions to obtain a new function; as long as \( g(z) \) is in the domain of \( f \), we can form \( f(g(z)) \), which is denoted \( f \circ g \). One then has \( f \circ g \) defined on some domain (the intersection of the domain of \( f \) with the range of \( g \)), and we think of \( f \circ g \) as a kind of multiplication of functions, one that is very noncommutative, but one that has an identity element \( \iota \). What do we mean by an identity element? The number 1 is called the multiplicative identity because \( 1 \cdot x = x \cdot 1 = x \) for all numbers \( x \). The number 0 is called the additive identity because \( 0 + x = x + 0 = x \) for all numbers \( x \). The identity function (there are actually many
of them!} satisfies \( \iota \circ f = f \) and \( f \circ \iota = f \), whenever those compositions are defined. Saying that \( \frac{1}{3} \) is the multiplicative inverse of \( \frac{1}{3} \) expresses the relationship \( (\frac{1}{3})(3) = 1 \). Saying that \( 5 \) is the additive inverse of \(-5\) expresses \(-5 + 5 + (-5) = 0\). The pair of functions \( f \) and \( g \) are inverses exactly when \( f \circ g = \iota \) and \( \iota = g \circ f \) (notice that the same symbol \( \iota \) is used to represent two possibly different functions).

**Problem 13** Give an example of an invertible function \( f : N \to N \) besides \( \iota \), and say what the inverse function is.

### 1.2 Partially defined functions

Sometimes only the rule of a function is given, so that the function is only partially defined. We need the reader to become adept at interpreting partial information about the domain and codomain by counting the number of coordinates that appear in an input or an output, and that is the purpose of the following definitions and exercises. When a rule is given via an equation, such as

\[
f(x, y) = x^2 + y,
\]

there is implicit information about the domain, namely that it is a set consisting of ordered pairs. In this case we will say that the **implicit domain** is \( R^2 \). More generally, if the inputs have \( m \) coordinates, we will say that the **implicit domain** is the subset of \( R^m \) for which the formula makes sense, and if the outputs have \( n \) coordinates, we will say that \( R^n \) is the **implicit codomain**. Thus our function \( f \) above has implicit domain \( R^2 \) and implicit codomain \( R^3 \).

**Example 7**

Determine the implicit domain, implicit codomain, and the range of the function given by

\[
f(x, y, z) = (x^2, xy, y^2, 0).
\]

The implicit domain is suggested by the number of coordinates of an input of \( f \); for any element \((x, y, z)\) of \( R^3 \), it makes sense to apply \( f \) to this vector, so the implicit domain is \( R^3 \). The implicit codomain is suggested by the number of coordinates in an output; the vectors \((x^2, xy, y^2, 0)\) are elements of \( R^4 \), so \( R^4 \) is the implicit codomain. The range of \( f \) is defined to be

\[
\{f(x, y, z) : (x, y, z) \in R^3\}
\]

and the fastest way to display the range is to simply write

\[
\{(x^2, xy, y^2, 0) : (x, y, z) \in R^3\}.
\]

In this particular instance, the range is a two dimensional surface in four dimensional space.

**Example 8**

Determine the implicit domain, implicit codomain, and range of the function defined by

\[
f(x, y) = (\sqrt{xy}, y^2, \sqrt{x}).
\]

This time the inputs are ordered pairs, and those folks who are in the habit of dealing with real numbers will recognize not every ordered pair makes sense in the formula. Since a square root is taken of \( x \), we would need to require \( x \geq 0 \), and \( \sqrt{xy} \) forces us to require that \( y \geq 0 \). Thus the implicit domain is the first quadrant in \( R^2 \), and the implicit codomain is \( R^3 \). The range is

\[
\{(\sqrt{xy}, y^2, \sqrt{x}) : x \geq 0, y \geq 0\}.
\]
Of course, we might equally well say that the implicit domain could be all of \( C^2 \), while the implicit codomain is \( C^3 \). If you give this answer, then the correct answer for the range would be

\[
\{(\sqrt{x^2}, y^2, \sqrt{x}) : x \in C, y \in C\},
\]

which is what people call a two dimensional complex manifold inside \( C^3 \). We could get carried away with this issue, giving further possible solutions involving domains where the algebra defining \( f \) makes sense. For the sake of simplicity, let us agree to stick with the sets \( R^n \) for the time being, and let us also agree not to make a distinction between an ordered pair (or n-tuple) written horizontally or vertically. We will soon have good reason to write vectors vertically.

**Example 9**

Determine the implicit domain, implicit codomain, and range of the function defined by

\[
f(\begin{pmatrix} x \\ y \\ z \end{pmatrix}) = \begin{pmatrix} x + y + z \\ x + y + z \end{pmatrix}.
\]

Looking inside the parenthesis in the expression

\[
f(\begin{pmatrix} x \\ y \\ z \end{pmatrix})
\]

reveals that the implicit domain is \( R^3 \), while counting the number of coordinates in

\[
\begin{pmatrix} x + y + z \\ x + y + z \end{pmatrix} \quad \text{First coordinate} \quad \text{Second coordinate}
\]

reveals that the implicit codomain is \( R^2 \). The range is the line \( \{(x, x) : x \in R\} \) in \( R^2 \), i.e. the line in \( R^2 \) whose equation is \( y = x \).

**Example 10**

Find the implicit domain and range of \( f(x, y) = (x^\frac{1}{2}, y^\frac{1}{2}) \).

Once again, the vectors being input into \( f \) are elements of \( R^2 \), but not every vector in \( R^2 \) is an allowable input; for example \((-1, 1)\) can not be input into \( f \). The implicit domain is

\[
\{(x, y) : x \geq 0, y \in R\},
\]

which is the right half plane in \( R^2 \). The range is the set

\[
\{(x^\frac{1}{2}, y^\frac{1}{2}) : x \geq 0, y \in R\},
\]

which is again the right half plane.

**Problem 14** For each of the following, state the implicit domain, implicit codomain, and range of the function.

1. \( f(x, y) = xy \)
2. \( f(x, y) = (5, 5) \)
3. \( f(\begin{pmatrix} x \\ y \\ z \end{pmatrix}) = \begin{pmatrix} x \\ y^2 \\ z^3 \end{pmatrix} \)
Problem 15 For each of the following, state the implicit domain and describe the range geometrically.

1. \( f(x, y) = (x^2, 2x^2) \)  
2. \( f(x, y) = x^2y^2 \)  
3. \( f\left(\frac{x}{y}\right) = \left(\frac{x+y}{x-y}\right) \)

1.3 Vectors and Coordinates

A free vector in space is a directed line segment. Two free vectors are equivalent if they have the same length and direction. A vector in space is a given specification of the pair direction–magnitude. Given any vector \( \mathbf{v} \), there are infinitely many free vectors with the same direction–magnitude as \( \mathbf{v} \), and we will feel free to select any one of these as a representative of \( \mathbf{v} \).

In the mathematical world, it is celebrated when two disparate mathematical areas are interconnected. Geometry meets algebra when Cartesian coordinates are introduced. Free vectors are geometric objects, but with the introduction of a coordinate system, we may associate each free vector with an algebraic object. If we are given a free vector, we locate the equivalent free vector whose tail lies at the origin and obtain the cartesian coordinates of the point located at its tip. Any two equivalent free vectors are associated with the same tuple of cartesian coordinates, so this association establishes a one-to-one correspondence between the set of cartesian coordinates and the set of all vectors. This gives us a dual way to think of vectors, geometric and algebraic, and we intend to pass between the two viewpoints freely. It is common practice to label a free vector with a symbol like ‘\( \mathbf{u} \)’, when in fact \( \mathbf{u} \) is intended to represent the vector, not the free vector, and any other arrow with the same direction and magnitude will be thought of as the same as \( \mathbf{u} \), as will the corresponding cartesian coordinate.

Example 11 To find the coordinates to associate with a free vector \( \mathbf{u} \), we locate the equivalent free vector whose tail lies at the origin, then we read the coordinates at its arrow tip. In Fig. 2 we estimate that \( \mathbf{u} \) is associated with the coordinates \((2, 2)\).

![Fig. 2](image)

Given a tuple’s cartesian coordinates, the process of identifying which free vectors to associate with that tuple is the reverse of what we just described: draw the vector whose tail is at the origin and whose tip lies at the point with those coordinates. Every free vector with the same length and direction is then associated with that tuple.

Problem 16 Estimate the coordinates associated with the arrow \( \mathbf{v} \) drawn in Fig. 3.

![Fig. 3](image)

There are free vectors that have no length and, when drawn, look like a dot. These lengthless free vectors represent the zero vector, denoted \( \mathbf{0} \), which is given coordinates \((0, 0)\). Vectors may be multiplied (scaled) by numbers (scalars): if \( \mathbf{v} \) is a given vector, then \( 2\mathbf{v} \) is the vector with the same direction and double the length of \( \mathbf{v} \). Two vectors \( \mathbf{v} \) and \( \mathbf{w} \) might be related algebraically with an equation \( \mathbf{w} = \alpha\mathbf{v} \). The corresponding geometric picture looks like two parallel arrows, with the arrow tips pointing in the same direction exactly when \( \alpha \) is positive. We will refer to \( \mathbf{v} \) and \( \mathbf{w} \) as a dependent pair in this case. Notice that \( \mathbf{v} \) and \( \mathbf{0} \) form a dependent pair for any vector \( \mathbf{v} \).

Problem 17 Draw two arrows representing vectors \( \mathbf{u} \) and \( \mathbf{w} \) when \( \mathbf{w} = -2\mathbf{u} \).

![Fig. 4](image)

When two vectors fail to be dependent, the pair of vectors is said to be independent. We visualize two independent vectors by drawing their representing arrows with joined tails. Two such arrows, being non-parallel, determine a parallelogram in a plane. The parallelogram rule for addition of vectors says that the diagonals of this parallelogram should represent the sum and difference of the two vectors.
Problem 18 Show that the parallelogram rule for adding vectors agrees with coordinate addition: if \( \mathbf{u} \) and \( \mathbf{v} \) have corresponding coordinates \((x_1, y_1)\) and \((x_2, y_2)\), show that the coordinates of \( \mathbf{u} + \mathbf{v} \) are \((x_1 + x_2, y_1 + y_2)\).

Another visualization of vector addition looks a bit like a snake. Start with an arrow \( \mathbf{u} \) and add to the tip of \( \mathbf{u} \) the tail of a second arrow \( \mathbf{v} \). The arrow whose tail touches the tail of \( \mathbf{u} \) and whose tip coincides with the tip of \( \mathbf{v} \) represents \( \mathbf{u} + \mathbf{v} \). If we tag the tail of a third vector \( \mathbf{w} \) to the tip of \( \mathbf{v} \), we get a picture like that in Fig. 4 depicting the sum \( \mathbf{u} + \mathbf{v} + \mathbf{w} \). Notice that \( \mathbf{u} \) and \( \mathbf{v} \) in Fig. 4 are equivalent to those in Fig. 5, and \( \mathbf{w} \) is equivalent to \( \mathbf{u} + \mathbf{v} \), so \( \mathbf{u} + \mathbf{v} + \mathbf{w} \) should be equivalent to \( 2(\mathbf{u} + \mathbf{v}) \), which it is.

A linear combination of \( \mathbf{u} \) and \( \mathbf{v} \) is a vector of the form \( \alpha \mathbf{u} + \beta \mathbf{v} \), with \( \alpha \) and \( \beta \) scalars. If \( \mathbf{u} \) and \( \mathbf{v} \) are dependent, then the result is a third vector parallel to both. If these vectors are independent, then every vector in the plane they determine can be written as a linear combination of \( \mathbf{u} \) and \( \mathbf{v} \). The span of \( \mathbf{u} \) and \( \mathbf{v} \) is defined to be the set of all linear combinations of the two vectors, so if the vectors are independent they span a plane and if the vectors are dependent (and not both zero) they span a line.

Problem 19 Draw a picture that illustrates the span of the vectors \( \mathbf{v} \) and \( \mathbf{u} \) when their corresponding coordinates are \((0, 2)\) and \((0, 5)\), respectively.

Problem 20 With \( \mathbf{u} \), \( \mathbf{v} \), and \( \mathbf{w} \) as drawn in Fig. 6, find good estimates for scalars \( \alpha \) and \( \beta \) so that \( \mathbf{w} = \alpha \mathbf{u} + \beta \mathbf{v} \).

The definitions of independent, linear combination, and span may be generalized so that they apply to a set of vectors with any number of elements. Assume \( \mathcal{S} \) is a set of vectors. A linear combination of vectors from \( \mathcal{S} \) is a vector of the form \( \alpha_1 \mathbf{v}_1 + \ldots + \alpha_n \mathbf{v}_n \), where each vector \( \mathbf{v}_i \) (for \( i = 1, \ldots, n \)) is in the set \( \mathcal{S} \) and each \( \alpha_i \) is a scalar. The length of the sum (represented by the integer \( n \)) is finite, but otherwise not restricted. The span of \( \mathcal{S} \) is the set of all possible linear combinations of vectors from \( \mathcal{S} \). When we call a set a span, we mean that it is the span of \( \mathcal{S} \) for some set \( \mathcal{S} \). We say that \( \mathcal{S} \) is independent when there exists only one way to obtain the zero vector as a linear combination (which is by using all zeros for scalars). Thus \( \mathcal{S} \) is independent exactly when the following is true: if \( \mathbf{v}_i \in \mathcal{S} \) and \( \alpha_1 \mathbf{v}_1 + \ldots + \alpha_n \mathbf{v}_n = \mathbf{0} \), then \( \alpha_1 = \ldots = \alpha_n = 0 \). If \( \mathcal{S} \) is not independent we will say it is dependent. A set is dependent exactly when one of the vectors in the set is a linear combination of other vectors in the set.

Problem 21 Give an example of two sets of vectors \( \mathcal{S}_1 \) and \( \mathcal{S}_2 \), where \( \mathcal{S}_1 \) contains only a single vector, \( \mathcal{S}_2 \) contains three distinct vectors, and the span of \( \mathcal{S}_1 \) is the same as the span of \( \mathcal{S}_2 \).

For a set of vectors in \( \mathbb{R}^3 \), notice that the span has a very particular structure; the span of the set can be one of four things. It could be all of \( \mathbb{R}^3 \), it could be a plane containing the zero vector, it could be a line containing the zero vector, or it could be the set that contains only the zero vector. Thus, when we refer to a set as a span, we have greatly limited what kind of set it can be. These special subsets that may be obtained as a span of vectors are called subspaces.

Example 12 Consider the sets \( \mathcal{S}_1 = \{(1, 1, 1), (0, 1, 0)\} \) and \( \mathcal{S}_2 = \{(1, 2, 1), (1, 0, 1)\} \). Notice that each vector in \( \mathcal{S}_2 \) is in the span of \( \mathcal{S}_1 \); the equality

\[
(1, 0, 1) = (1, 1, 1) - (0, 1, 0)
\]

proves this assertion for the second vector in \( \mathcal{S}_2 \). It follows that the span of \( \mathcal{S}_2 \) is contained in the span of \( \mathcal{S}_1 \). Since both vectors in \( \mathcal{S}_1 \) are seen to be in the span of \( \mathcal{S}_2 \), we conclude that \( \mathcal{S}_1 \) and \( \mathcal{S}_2 \) have the same span, which is a plane in \( \mathbb{R}^3 \) containing the origin.
The span of one set may also be the span of many other sets. For example, if $S$ is dependent, so that some vector $\mathbf{v}$ in $S$ is a linear combination of other vectors in $S$, then we can remove $\mathbf{v}$ from $S$ and the resulting smaller set will have the same span as $S$. This process of removing vectors will ultimately end with an independent subset of $S$ with the same span as $S$. One may think of independent sets as minimal spanning sets: if $S$ is independent and a vector is removed from $S$, then the resulting smaller set will have a distinctly smaller span than $S$. We refer to an independent set as a \textbf{basis} for its span. There are always many different bases for a particular subspace, but it is a fact that each of those bases has the same number of elements. This number of elements is called the \textbf{dimension} of the subspace. A basis for a line containing the zero vector has one vector. A basis for a plane containing the zero vector has two vectors. A basis for $\mathbb{R}^3$ has three vectors.

\textbf{Problem 22} Give an example of two sets of vectors $S_1$ and $S_2$ with the following properties: both $S_1$ and $S_2$ are independent sets, both contain exactly two vectors, and their spans intersect in a line.

1.4 Visualizing transformations of the plane

Given a function $f : \mathbb{R} \rightarrow \mathbb{R}$, we generally draw a graph of $f$ in order to obtain a visualization of the global behaviour of $f$. The graph is drawn in $\mathbb{R}^2$ because the definition of its graph is \{$(x, f(x)) : x \in \mathbb{R}$\}. For a function $f : \mathbb{R}^2 \rightarrow \mathbb{R}$, the definition of the graph is \{$(\mathbf{v}, f(\mathbf{v})) : \mathbf{v} \in \mathbb{R}^2$\}, and each pair $(\mathbf{v}, f(\mathbf{v}))$ actually gives rise to three real coordinates, two coordinates for $\mathbf{v}$ and one coordinate for $f(\mathbf{v})$. The graph of this function can then be drawn as a subset of $\mathbb{R}^3$. If $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ and $\mathbf{v} \in \mathbb{R}^2$, then $(\mathbf{v}, f(\mathbf{v}))$ has four real coordinates, two coordinates for $\mathbf{v}$ and two for $f(\mathbf{v})$, so its graph is actually a subset of $\mathbb{R}^4$, which can not be drawn. For this reason, we need to introduce alternative ways of visualizing the action of such functions.

Assume that $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a function. The most straightforward way to visualize the action of $f$ is to draw a picture in the plane, then see how the picture distorts upon the application of $f$. If the baby’s ear in figure 7 lies at the coordinate $(1,1)$, then it is redrawn at coordinate $f(1,1)$, and likewise for every other pixel in the picture. The picture in figure 7, besides being the image before transforming, is also the result after applying the function $f(x,y) = (x, y)$. This is the \textbf{identity function} on $\mathbb{R}^2$, the function that puts every pixel back exactly where it was found.

Consider the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$f(x,y) = (x, \frac{1}{2}y).$$

The action of this function is to compress the points of the plane vertically toward the $x$-axis. The baby’s ear is moved down so that it’s new height above the axis is exactly half of what it was.

Let’s look at the function $f : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ defined by

$$f(x,y) = (\frac{3}{4}x + \frac{1}{4}y, \frac{1}{4}x + \frac{3}{4}y),$$

which is the function responsible for the distortion of the image drawn in figure 9. The equation of this function bears little resemblance to the one in the preceding paragraph, but the actions of the two are very similar. In fact, the illustration in figure 9 was prepared by rotating the original image $45^\circ$ clockwise, applying the function in the preceding paragraph, then rotating $45^\circ$ back. In figure 9, the line $y = x$ plays exactly the same role that the $x$-axis does in in figure 8; every pixel in the illustration of figure 9 was moved towards the line $y = x$ so that its distance to this line ends up being exactly half of what it was originally.

The functions we have introduced to illustrate the effect of transforming a photograph are examples of \textit{linear transformations}, which also happens to be the title of this book. The trick to understanding this type of function is to find the subspaces that are transformed back into themselves, what we like to call \textbf{invariant subspaces}. Recall that a subspace is the result of spanning some set, and a subspace of $\mathbb{R}^2$ is either a line containing the origin (a \textbf{nontrivial subspace}), or it
Chapter 1  •  Functions

is \{0\} or \(R^2\) itself (a trivial subspace). There are exactly two nontrivial invariant subspaces of the function illustrated in figure 8, the \(x\)-axis and the \(y\)-axis. Likewise, there are exactly two nontrivial invariant subspaces of the function illustrated in figure 9, the line \(y = x\) and the line \(y = -x\).

**Problem 23** Assume that \((x, y)\) is a point that is on neither the \(x\)-axis nor the \(y\)-axis. With \(f\) as in figure 8, show that \(f(x, y)\) does not lie on the span of \\{(x, y)\}\. 

**Problem 24** Assume that \((x, y)\) is a point that is on neither the line \(y = x\) nor the line \(y = -x\). With \(f\) as in figure 9, show that \(f(x, y)\) does not lie on the span of \\{(x, y)\}\. 

The method of visualizing the effect of a function on a photograph is helpful, but there are limitations; it is very hard to see the invariant subspaces in general, and some very simple functions cause the photograph to literally disappear. The projection \(f : R^2 \rightarrow R^2\) onto the \(x\)-axis is defined by the formula

\[ f(x, y) = (x, 0), \]

and the effect of this function is to take each point in the plane to the point on the \(x\)-axis closest to it. If this function were to act on all the pixels in our photograph, the result would be no picture. What is actually illustrated in figure 10 is the result of using a function like \(f(x, y) = (x, \epsilon y)\), with \(\epsilon\) very small.

### 1.5 Vector fields

Physicists use a method of visualizing vector functions that appeals strongly to our physical intuition. To illustrate the method with an example, suppose a two dimensional model is made to give a mathematical description of a fluid whirlpool. The origin is placed at the center of the whirlpool, and a Cartesian coordinate system is introduced. At each point \((x, y)\) the fluid has a particular velocity, which we denote \(f(x, y)\), and this gives rise to a function \(f : R^2 \rightarrow R^2\). One then creates the vector field for \(f\) by drawing an arrow representing \(f(x, y)\) with its tail at the point \((x, y)\). If this is done for too many points the result is a mess, but by selecting a few points judiciously one might arrive at a picture that gives the feeling of a swirling fluid, like that pictured in figure 11. A typical arrow indicates where the point at its tail is moving; it moves in the direction of the arrow, and its speed is proportional to the length of the arrow. This vector field certainly evokes the sense of a swirl, but anyone who has experienced the flush of a toilet knows that the water near the center is moving faster than the water at the edges, and our vector field is saying just the opposite. The truth is, we had a hard time drawing the longest arrows near the center of the picture, which points to a deficiency of the method; vector fields are frequently very hard to draw.

**Problem 25** Draw a vector field that represents a whirlpool better than the one in fig. 11

**Problem 26** Give an example of a physical object that is well represented by the vector field in fig. 11

Fortunately for us, vector fields come out quite nicely for linear transformations, and they make the invariant subspaces nicely visible. We draw the vector field by chaining vectors head to tail and coloring them from darkest to lightest. In this way we do not need to draw the arrow tips, since the arrows point in the direction from darkest to lightest. Study the example given in figure 12 of the vector field for the identity function. The white disc in the middle is the unit disc, i.e. it consists of all vectors of \(R^2\) whose distance to the origin is at most one. The black lines correspond to the vectors \(f(x, y)\) whose tails are at \((x, y)\) for 32 of the \((x, y)\) on the unit circle. Since \(f\) is the identity function, we see \(f(x, y) = (x, y)\) and notice that each of the black vectors actually represents the corresponding \((x, y)\) itself. The dark green lines are vector representatives of \(f(x, y)\) when \((x, y)\) lies at the tip of the black line, and the light green lines are the representatives of \(f(x, y)\) when \((x, y)\) lies at the tip of the corresponding dark green line. Keep in mind, we know what the tips and tails are by thinking ‘darkest to lightest’.
Problem 27  Draw a vector field for the function \( f(x, y) = \left( \frac{1}{2} x, \frac{1}{2} y \right) \).

Imagine drawing the coordinate axes in figure 12, with the origin at the center of the white disc, and the \( x \) and \( y \) axes in their usual horizontal and vertical positions. We see that four of the vector triples (represented by the black, dark green, and light green lines) are already in line with our axes, and moreover, every other vector triple seems to be in line with a subspace, i.e. with a line containing the origin. Any nonzero \((x, y) \in \mathbb{R}^2\) determines a nontrivial subspace, namely, the span of \( \{(x, y)\} \), and this subspace looks like the line through \((x, y)\) and the origin. The vector \( f(x, y) \) is in this subspace exactly when any one of its free vector representatives is parallel to this line, so when the tail is placed at \((x, y)\) we see the free vector cover a portion of the line. When \( f \) is the identity function, this will happen for every \((x, y) \in \mathbb{R}^2\). When \( f \) is some other linear function, this will happen for just those \((x, y) \in \mathbb{R}^2\) whose spans are invariant subspaces.

Problem 28  Draw a vector field for a function that has exactly one invariant subspace.

Let us return now to the function that appeared in figure 8, which was

\[ f : \mathbb{R}^2 \to \mathbb{R}^2 \text{ with } f(x, y) = (x, \frac{1}{2} y). \]

Example 13  With the vector field of \( f \) drawn in figure 13, find the two invariant subspaces of \( f \). Verify the invariance by algebraically proving that \( f(\mathbf{v}) \) is a multiple of \( \mathbf{v} \) for vectors on these lines.

We see that the \( x \)-axis and \( y \)-axis are where the vector triples line up, so these are the invariant subspaces. We check half of this answer by taking a typical vector on the \( x \)-axis, applying \( f \) to it, and seeing if the result is still on the \( x \)-axis. The typical vector on the \( x \)-axis looks like \( \mathbf{v} = (x, 0) \), and

\[ f(\mathbf{v}) = f(x, 0) = (x, \frac{1}{2} 0) = (x, 0) = \mathbf{v}, \]

so the vectors \( \mathbf{v} \) on the \( x \)-axis are actually taken to themselves.

We check the second half of the answer by taking \( \mathbf{v} = (0, y) \) and making sure \( f(\mathbf{v}) \) is still on the \( y \)-axis. We compute

\[ f(\mathbf{v}) = f(0, y) = (0, \frac{1}{2} y), \]

so such vectors are indeed taken to multiples of themselves; namely \( f(\mathbf{v}) = \frac{1}{2} \mathbf{v} \).

Example 14  Assume that \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) is defined by

\[ f(x, y) = \left( \frac{3}{4} x - \frac{1}{4} y, -\frac{1}{4} x + \frac{3}{4} y \right). \]

Determine the invariant subspaces of \( f \) from the vector field drawn in figure 14. Verify the invariance by demonstrating that \( f(\mathbf{v}) \) is a multiple of \( \mathbf{v} \) for vectors on these lines.

The vector triples seem to line up along the lines \( y = x \) and \( y = -x \). Starting with a vector \( \mathbf{v} \) on the line \( y = x \), we write \( \mathbf{v} = (x, x) \) and compute

\[ f(\mathbf{v}) = f(x, x) = \left( \frac{3}{4} x - \frac{1}{4} x, -\frac{1}{4} x + \frac{3}{4} x \right) = \frac{1}{2} (x, x) = \frac{1}{2} \mathbf{v}, \]

so these vectors are taken to multiples of themselves. Now we check \( \mathbf{v} = (x, -x) \) by computing

\[ f(\mathbf{v}) = f(x, -x) = \left( \frac{3}{4} x + \frac{1}{4} x, -\frac{1}{4} x - \frac{3}{4} x \right) = (x, -x) = \mathbf{v}, \]
and we see these vectors are also taken to multiples of themselves, that multiple being 1.

**Problem 29** Assume that \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is defined by

\[
 f(x, y) = \left( \frac{3}{4} x + \frac{1}{4} y, \frac{1}{4} x + \frac{3}{4} y \right).
\]

Determine the invariant subspaces of \( f \) from the vector field drawn in figure 15. Verify the invariance by demonstrating that \( f(\mathbf{v}) \) is a multiple of \( \mathbf{v} \) for vectors on these lines.

The reader should carefully compare the vector fields that appear in figures 13, 14, and 15; one of them can be obtained from another by a rotation. One of the major themes of this book is how one may group similar transformations together to distinguish them from other groups. Each of these three transformations will be grouped together because the only difference in their action is the somewhat random choice of how we draw our coordinate system. Each one scales one coordinate by a half and leaves the other coordinate alone. If some dizzy person were to draw the positive \( x \)-axis in figure 15 off in the northeast direction, and the positive \( y \)-axis is drawn in the northwest direction, then we would be looking at the vector field of the function \( f(x, y) = (x, 5y) \). In this way we end up distinguishing between groups of transformations instead of individual ones. Finding the invariant subspaces is the crucial step in this process, and the importance of the vectors on these invariant lines leads us to give them a special name; an eigenvector is a nonzero vector on an invariant line. Such vectors \( \mathbf{v} \) may be characterized by the fact that the transformation takes \( \mathbf{v} \) to some scalar multiple of \( \mathbf{v} \); that scalar multiple is called the eigenvalue corresponding to the eigenvector.

**Problem 30** In figure 14, indicate how the \( x \) and \( y \) axes should be drawn so that the vector field is the one for the function \( f(x, y) = (x, 5y) \).

**Exercises**

**Pictures** Solutions to earlier exercises are often useful when doing the later ones. Draw a picture depicting the following sets.

1. The set \( A \equiv \{ \alpha \mathbf{v} : \alpha \in \mathbb{R} \} \) as a subspace of \( \mathbb{R}^2 \) when \( \mathbf{v} = (1, 1) \).
2. The set \( B \equiv \{ \alpha \mathbf{v} : \alpha \in \mathbb{R} \} \) as a subspace of \( \mathbb{R}^3 \) when \( \mathbf{v} = (1, 1, 1) \).
3. The span of \( C \equiv \{ (1, 0, 0), (1, 1, 1) \} \) as a subspace of \( \mathbb{R}^3 \).
4. The set \( D \equiv \{ (1, 2) + \alpha (1, 1) : \alpha \in \mathbb{R} \} \) as a subset of \( \mathbb{R}^2 \).
5. The set \( E \equiv \{ \alpha (1, 2) + (1 - \alpha) (2, 3) : 0 \leq \alpha \leq 1 \} \) as a subset of \( \mathbb{R}^2 \).
6. The set \( F \equiv \{ \alpha (0, 0) + \beta (1, 1) + \gamma (0, 1) : 0 \leq \alpha, \beta, \gamma \leq 1, \alpha + \beta + \gamma = 1 \} \) as a subset of \( \mathbb{R}^2 \).
7. The set \( A \cap D \) as a subset of \( \mathbb{R}^2 \).
8. The set \( A \cap F \) as a subset of \( \mathbb{R}^2 \).
9. The set \( B \cap C \) as a subset of \( \mathbb{R}^3 \).
10. The set \( A \cap B \).

**Equivalence relations** This is the first of a series of exercises on the topic of equivalence relations. There is no need to know what an equivalence relation is in order to ponder the following exercises; they are intended to gently lead the reader towards an understanding.

11. In the vector field drawn in fig. 12, describe a way to separate the 96 arrows drawn into 32 groups of 3. This is an example of a partition of the 96 arrows.
12. Describe how one might partition the set of all free vectors so that the groups formed correspond naturally to the vectors. (The answer appears in the discussion at the beginning of section 3 on page 7).
13. Describe how to partition \( \mathbb{R}^3 \) into planes. How might \( \mathbb{R}^3 \) be partitioned into lines?
14. If we think of the elements of \( \mathbb{R}^4 \) as being pairs of pairs of real numbers, then we’re really thinking of the elements as \( (\mathbf{u}, \mathbf{v}) \) with \( \mathbf{u}, \mathbf{v} \in \mathbb{R}^2 \). Describe how one might partition \( \mathbb{R}^4 \) into pieces that look like \( \mathbb{R}^2 \).

**Examples** Give an example of...

15. a subspace with only one vector. Is there a subspace that has two vectors?
16. an independent set with 4 vectors. Can an independent set have infinitely many vectors?
17. a subset of \( \mathbb{R}^2 \) that is not a subspace. Explain your reasoning.
18. a function \( f : Z \rightarrow N \) that is onto.
19. a function \( f : Z \rightarrow N \) that is one-to-one.
20. a function \( f : \mathbb{R} \rightarrow \mathbb{R} \) that is one-to-one but not onto.

21. a function \( f : \mathbb{R} \rightarrow \mathbb{R} \) that is onto but not one-to-one.

22. a function \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) that is onto.

23. a function \( f : \mathbb{R}^2 \rightarrow \mathbb{R} \) that is one-to-one.

**Counterexamples** Use a counterexample to show that the following statements are false.

24. The length of \( \mathbf{u} + \mathbf{v} \) is the length of \( \mathbf{u} \) plus the length of \( \mathbf{v} \).

25. The length of the complex number \( \alpha \) is \( \sqrt{\alpha^2} \).

26. The codomain of a function is the same as the range of the function.

27. The sum of two onto functions is onto.

28. The square of an invertible function is invertible.

29. In \( \mathbb{R}^2 \), a set with two vectors in it is a basis.

30. If \( \mathcal{B} \) is a basis of \( \mathbb{R}^3 \), and \( \mathcal{C} \) is a three element set containing two of the vectors of \( \mathcal{B} \) but a different third vector, then \( \mathcal{C} \) is not a basis of \( \mathbb{R}^3 \).

31. The union of two independent sets is independent.

32. The intersection of two spanning sets is a spanning set.

33. The complement of a spanning set is an independent set.

**Et cetera**

34. Draw a vector field for \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) when \( f(x, y) = (0, y) \).

35. Draw a vector field for \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) when \( f(x, y) = -(x, y) \).

36. Draw a vector field for \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) when \( f(x, y) = (y, x) \).

37. Draw a vector field for \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) when \( f(x, y) = (0, 0) \).

38. Draw a vector field for \( f : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) when \( f(x, y) = (x, x + y) \).
Chapter 1  ●  Functions

More Exercises

1. Coordinate functions...
2. Equivalence Relations
3. Devise a way to associate coordinates with free vectors in $R^2$ so that we get analogues of
4. Draw a picture depicting $\{\alpha \mathbf{v} : \alpha \in R\}$ when $\mathbf{v}$ is the vector with corresponding coordinates $(1, 2)$
5. Assume that a vector $\mathbf{v}$ is associated with the coordinates $(a, b) \in R^2$. Use the Pythagorean theorem to show that, relative to the scale chosen for the coordinate system, the length of $\mathbf{v}$ is $\sqrt{a^2 + b^2}$.
6. In the paragraph opposite fig. we assert that $\mathbf{u} + \mathbf{v} + \mathbf{w} = 2(\mathbf{u} + \mathbf{v})$; give a geometric proof.
7. Free vectors are really vectors after all, but vectors of a higher dimension. Consider the set $\mathcal{V}$ of all free vectors in $R^2$. Show that the free vectors of $\mathcal{V}$ can be put in one-to-one correspondence with the four-tuples in $R^4$.
8. Using your pairing of the set of free vectors with $R^4$, find the subset of $R^4$ that corresponds to the set of free zero vectors (see the discussion in the paragraph next to Fig. 3).
9. Using your pairing of the set of free vectors with $R^4$, find the subset of $R^4$ that corresponds to the set of free zero vectors that represent the vector with coordinates $(0, 1)$.
10. Give a geometric verbal description of the two preceding answers, and emphasize how the two sets are related to each other.
Chapter 2

Coordinates

2.1 A generalization

Cartesian coordinates are familiar to people who have ever graphed a line like $y = 2x$. An easy way to graph that line is to find two points $(x_1, y_1)$ and $(x_2, y_2)$ that satisfy the equation $y = 2x$, like $(0, 0)$ and $(1, 2)$, then draw a straight line through those two points. We are taught to first draw a vertical and a horizontal line, which serve as the coordinate axes, create a scale on each axis, then use that scale to locate $(1, 2)$ in the upper right quadrant. The line is then drawn through the point $(1, 2)$ and the origin $(0, 0)$, which is where the coordinate axes intersect. Any expression like $(x, y)$ is called an ordered pair, and we are trained to associate the first number $x$ with a position on the horizontal axis, and to associate the second number $y$ with a position on the vertical axis. The two line segments joining the origin to the $x$ and $y$ positions suggests a rectangle, and the edge of that rectangle that is opposite the origin is what we are trained to associate with the ordered pair $w = (x, y)$ (Figure 1).

![Figure 1: The point $w = (x, y)$](image1)

> Fig. 1: The point $w = (x, y)$

It is common to see a different notation used to describe the same thing as $(x, y)$. If we let $i$ denote $(1, 0)$, and let $j$ denote $(0, 1)$, then $xi + yj$ denotes the same thing as $(x, y)$. Regardless of the notation used, the two numbers $x$ and $y$ are called the coordinates of the vector associated with $(x, y) = xi + yj$. The first direction a generalization takes is to allow our vectors to have more than two coordinates, so we consider spaces whose elements are objects of the form

$$(x_1, x_2, \ldots, x_n),$$

and call these objects $n$-tuples, or simply tuples. If we agree to let $e_i$ denote the tuple whose $i^{th}$ coordinate is one, and all other coordinates zero (for each $i = 1, 2, \ldots, n$), then we may use the alternate notation

$$(x_1, x_2, \ldots, x_n) = x_1e_1 + x_2e_2 + \ldots + x_ne_n.$$  

In the case of two coordinates, this sum is depicted in Figure 2 (compared with Figure 1).

![Figure 2: $w = xe_1 + ye_2$](image2)

> Fig. 2: $w = xe_1 + ye_2$

The second direction our generalization takes is to allow the vectors $e_i$ to be replaced with any linearly independent set, which we illustrate with two independent vectors $u$ and $v$ in the plane. In Figure 3 we see $u$ and $v$ drawn with their tails joined. The point where the tails join determines the origin of this coordinate system, and the two arrows determine the two axes. In order for us to assign first and second coordinates relative to these axes, we need to specify which of $u$ and $v$ is first and which is second; we write $\{u, v\}$ to indicate the choice of $u$ first and $v$ second, and we label this ordered set $B \equiv \{u, v\}$. Once this is done, we consider $(x, y)_B$ to be the vector $xu + yv$, and we call $x$ and $y$ the $B$-coordinates of the vector $w = xu + yv$. Every vector in the plane may be expressed in $B$-coordinates by determining how to write that vector as a linear combination of $u$ and $v$. In particular, the vectors $u$ and $v$ acquire the $B$-coordinates of $(1, 0)_B$ and $(0, 1)_B$ (respectively), and
their sum has the \( B \)-coordinates \((1, 1)_B\) (Figure 3). In case \( B = \{e_1, e_2\}\), i.e. the case in which the first vector in \( B \) is a horizontal vector of length one and the second vector is the vertical vector with length one, then the \( B \)-coordinates are the usual cartesian coordinates, which we call the **standard coordinates** (Figure 4). In all other cases we expect the \( B \)-coordinates to be quite different than their standard coordinates.

**Problem 1** If \( B = \{e_1, e_2\}\), give the \( B \)-coordinates of the vector whose cartesian coordinates are \((1, 2)\).

**Problem 2** If \( B = \{e_2, e_1\}\), give the \( B \)-coordinates of the vector whose cartesian coordinates are \((1, 2)\).

**Problem 3** Trace the image in Figure 3, label one of the arrow \( u \) and the other \( v \), then plot and label the vectors corresponding to the \( B \)-coordinates \((-1, -1)_B\), \((1, -1)_B\), and \((-1, 1)_B\).

**Problem 4** What happens to the coordinates assigned in Problem 3 if the arrow labeled \( u \) is relabeled \( v \) and the arrow labeled \( v \) is relabeled \( u \)?

**Problem 5** Trace the image in Figure 4 and label the arrows so that \( \{u, v\} \) corresponds to standard coordinates. Plot and label the points corresponding to the standard coordinates \((-1, -1)\), \((1, -1)\), and \((-1, 1)\).

**Problem 6** Combine the answers to problems 3 and 5 in a single drawing so that the origins of the two systems coincide.

### 2.2 Tayloring coordinates to the function

Frequently we are presented with a linear function given in terms of one coordinate system, for which we build a second, more appropriate coordinate system. When we need to simultaneously talk about two different coordinate systems, we will use two letters to denote the two ordered bases, \( B \) and \( C \) being good candidates, so if \( B = \{u, v\} \) and \( C = \{u', v'\} \), then \((3, -5)_C\) refers to the vector \(3\mathbf{u'} - 5\mathbf{v'}\), while \((3, -5)_B\) refers to the vector \(3\mathbf{u} - 5\mathbf{v}\). When no subscript appears on the coordinate tuple, we are to assume that standard coordinates are used. In the illustrations of our vector fields, where no coordinate axes are drawn, we are to assume that the center of the illustration represents the origin, and the standard coordinate axes intersect there, with the \( x \) axis horizontal and the \( y \) axis vertical.

**Example 1**

In Figure 5 we see the vector field for the function \( f : R^2 \rightarrow R^2 \) defined by

\[
f(x, y) = \left( \frac{5x}{12} - \frac{y}{12}, \frac{-x}{12} + \frac{5y}{12} \right).
\]

Find the invariant subspaces and the corresponding eigenvalues of \( f \), and use these to obtain a new and improved formula for \( f \).

We see the vectors align along the \( y = x \) line and the \( y = -x \) line, so these are the invariant subspaces. To see what the eigenvalues are, recall that these are the values by which the invariant subspaces are scaled, and also recall that the white disc in the center of the illustration has radius 1. By comparing this radius to the length of the black vectors on the line \( y = x \), we can determine the amount the vectors on this line are scaled by. A ruler reveals that these black vectors are \( 1/3 \) the length of the radius, so the eigenvalue corresponding to the line \( y = x \) is \( 1/3 \). A ruler measurement of the black vectors on the line \( y = -x \) shows their lengths to be half the radius of the white disc, so the eigenvalue corresponding to this line is \( 1/2 \). If we now let \( u \) and \( v \) be the northeast and northwest
(respectively) vectors drawn in Figure 6, and label $B = \{u, v\}$, the new formula for $f$ assumes the form $f(x, y)_B = (\frac{5}{2}x, \frac{5}{2}y)_B$.

**Problem 7**

Let $f : \mathbb{R}^2 \to \mathbb{R}^2$, defined by

$$f(x, y) = \left(\frac{5}{2}x + \frac{y}{2}, \frac{x}{2} + \frac{5}{2}y\right),$$

be the function whose vector field is drawn in Figure 7. Find the invariant subspaces and the corresponding eigenvalues of $f$, and use these to obtain a new and improved formula for $f$.

There are parts of this story we haven’t revealed to the reader yet. In Example 1, we can see that the new formula for $f$ works for the vectors on the two lines, i.e. for vectors of the form $(x, 0)_B$ and $(0, y)_B$, but it might be mysterious why the new formula works for every vector in the plane, i.e. for all $(x, y)_B$.

**Problem 8**

Speculate why the new formula for $f$ is valid for all vectors $(x, y)_B$ in the plane.

When we specify a function $f : \mathbb{R} \to \mathbb{R}$ by giving a formula like

$$f(x, y) = \left(\frac{5}{2}x + \frac{y}{2}, \frac{x}{2} + \frac{5}{2}y\right),$$

we are doing so relative to the standard coordinates that define $\mathbb{R}^2$; the set of all pairs of real numbers $(x, y)$ comes with the coordinates determined by the two vectors $e_1 = (1, 0)$ and $e_2 = (0, 1)$. When we seek to replace the standard coordinates with a coordinate system more suited to the function $f$, we may refer to the new pair of basis vectors in terms of their standard coordinates. Introducing $u = (1, 1)$ and $v = (-1, 1)$ in problem 7 gives us the appropriate basis $B = \{u, v\}$, and we now have the algebraic machinery needed to check our geometric estimates. Exactly what multiple of $u$ is $f(u)$? We compute

$$f(u) = f(1, 1) = \left(\frac{5}{2} + \frac{1}{2}, \frac{1}{2} + \frac{5}{2}\right) = (3, 3),$$

which answers the question; it is 3.

The vector field for the function $f(x, y) = (x, 0)$ appears in Figure 8; this function projects a point in the plane to the point on the $x$-axis closest to it. One of the invariant subspaces is visible, namely the $x$-axis, but the other invariant subspace is slightly obscured. The points on the $y$-axis are all taken to the origin, so the $y$-axis is an invariant line, but the vector field does not show any arrows lining up in that direction because those arrows have no length. An examination of this vector field shows the lengths of arrows diminishing the closer the arrow is to the $y$-axis, and the arrows disappear altogether at the $y$-axis. This is typical of a vector field of a projection onto a line. (Notice that the unit disc in the center appears as an ellipse; this is because of the different scales used on the two axes.)
Example 2

The vector field for the function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by
\[
f(x, y) = (0, y)
\]
is drawn in Figure 9. Find the standard coordinates of vectors \( \mathbf{u} \) and \( \mathbf{v} \) so that, relative to the basis \( \mathcal{B} = \{ \mathbf{u}, \mathbf{v} \} \), the new formula of \( f \) becomes \( f(x, y)_{\mathcal{B}} = (x, 0)_{\mathcal{B}} \). Verify your answer by computing \( f(\mathbf{u}) \) and \( f(\mathbf{v}) \) in the standard coordinates.

One solution is obtained by letting \( \mathbf{u} = (0, 1) \) and \( \mathbf{v} = (1, 0) \), which the reader is invited to verify. These problems do not have unique correct answers, however, and we will emphasize this by presenting the alternate solution \( \mathbf{u} = (0, -1) \) and \( \mathbf{v} = (4, 0) \). With this choice we see that
\[
f(\mathbf{u}) = f(0, -1) = (0, -1) \quad \text{and} \quad f(\mathbf{v}) = f(4, 0) = (0, 0),
\]
so \( f(\mathbf{u}) = \mathbf{u} \) and \( f(\mathbf{v}) = \mathbf{0} \). Relative to \( \mathcal{B} = \{ \mathbf{u}, \mathbf{v} \} \), the new formula for \( f \) is \( f(x, y)_{\mathcal{B}} = (x, 0)_{\mathcal{B}} \).

Problem 9

The vector field for the function \( f : \mathbb{R}^2 \to \mathbb{R}^2 \) defined by
\[
f(x, y) = \left( \frac{1}{2}x + \frac{1}{2}y, \frac{1}{2}x + \frac{1}{2}y \right)
\]
is drawn in Figure 10. Find the standard coordinates of vectors \( \mathbf{u} \) and \( \mathbf{v} \) so that, relative to the basis \( \mathcal{B} = \{ \mathbf{u}, \mathbf{v} \} \), the new formula of \( f \) becomes \( f(x, y)_{\mathcal{B}} = (x, 0)_{\mathcal{B}} \). Verify your answer by computing \( f(\mathbf{u}) \) and \( f(\mathbf{v}) \) in the standard coordinates.

Rotating the vector field in Figure 10 gives us the vector fields of infinitely more transformations that essentially all do the same thing; they all project onto a line.
2.3 Changing coordinates

We now need to confront the problem of how to pass between different coordinate systems. For the beginner, the most important thing to remember is that the two vectors that are determining a particular coordinate system are given the distinguished coordinates \((1,0)_\mathcal{B}\) and \((0,1)_\mathcal{B}\), after which all other vectors obtain their coordinates \((\alpha, \beta)_\mathcal{B} = \alpha(1,0)_\mathcal{B} + \beta(0,1)_\mathcal{B}\). If a second basis \(\mathcal{C}\) is in the discussion, then the vectors in that basis receive the distinguished coordinates \((1,0)_\mathcal{C}\) and \((0,1)_\mathcal{C}\), and if the \(\mathcal{B}\)-coordinates of these two vectors are also known, then it’s possible to pass back and forth between coordinate systems.

Example 3

In Figure 11 let us write \(\mathcal{B} = \{u, v\}\) and \(\mathcal{C} = \{u', v'\}\). Write \(u, v,\) and \(w\) in \(\mathcal{B}\)-coordinates and in \(\mathcal{C}\)-coordinates. Balance your coordinate estimates so that the sum \(u + v = w\) is accurate in both systems.

![Diagram](image)

The easy part of this problem is writing everything in \(\mathcal{B}\)-coordinate. Since \(u\) and \(v\) are the vectors in the set \(\mathcal{B}\) they get the coordinates \((1,0)_\mathcal{B}\) and \((0,1)_\mathcal{B}\), and since \(w\) is their sum, it gets the coordinates \((1,1)_\mathcal{B}\). Thus the equation \(u + v = w\) in \(\mathcal{B}\)-coordinates assumes the form

\[
(1,0)_\mathcal{B} + (0,1)_\mathcal{B} = (1,1)_\mathcal{B}.
\]

Now we write \(w\) in \(\mathcal{C}\)-coordinates; we are told that \(\mathcal{C}\) is the basis containing the green vectors, with \(u'\) first and \(v'\) second. Imagine extending the axes that these vectors determine, done so with the long dashed lines in Figure 11. From the tip of \(w\), we complete the parallelogram suggested by the coordinate axes, this completion drawn with short dotted lines in Figure 11. The \(\mathcal{C}\)-coordinates are determined from the sides of this parallelogram: we need to estimate \(\alpha\) and \(\beta\) so that \(\alpha u'\) is the vector determining the southeastern-most edge, and \(\beta v'\) is the vector determining the northwestern-most edge of the parallelogram. It looks like \(-1\) for \(\alpha\) and \(2\) for \(\beta\) work nicely, so the \(\mathcal{C}\)-coordinates are about \((-1,2)_\mathcal{C}\). Completing the parallelogram from the tip of \(u\) gives its coordinates, which appear to be about \(-1.2\) in the \(u'\) direction and perhaps \(.7\) in the \(v'\) direction; so our estimate for \(u\) is \((-1.2, 7)_\mathcal{C}\). In order for the coordinates to reflect the equation \(u + v = w\), we can compute

\[
v = w - u = (-1,2)_\mathcal{C} - (-1.2,.7)_\mathcal{C} = (2,1.3)_\mathcal{C}
\]

to obtain an estimate for the \(\mathcal{C}\)-coordinates of \(v\). To check for error, we now look at the picture in Figure 11 to see if these coordinates are reasonable; the \(.2\) seems really good, but the \(1.3\) appears quite big. We must have underestimated in the \(v'\) direction earlier in the process.

Problem 10

In Figure 12 let us write \(\mathcal{B} = \{u, v\}\) and \(\mathcal{C} = \{u', v'\}\). Write \(u, v,\) and \(w\) in \(\mathcal{B}\)-coordinates and in \(\mathcal{C}\)-coordinates. Balance your coordinate estimates so that the sum \(u + v = w\) is accurate in both systems.
Chapter 2  Coordinates

We could indicate which vectors are in a basis \( \mathcal{B} \) by using their standard coordinates. For example, we could define \( \mathcal{B} \) to be the basis \( \mathcal{B} = \{(1, 1), (0, 1)\} \), we then define the vector \( \mathbf{w} = (5, 8) \), and, letting \( \mathcal{C} \) be the standard basis, we could mimic the task in Problem 10. Bear in mind that writing \((1, 1)\) refers to a vector in its standard coordinates in \( \mathbb{R}^2 \), which by our choice of \( \mathcal{C} \) is, for the moment, the same as \((1, 1)_\mathcal{C}\). Tuples that carry no subscript means that standard coordinates are being used.

Example 4

In \( \mathbb{R}^2 \), let \( \mathcal{B} = \{(1, 1), (0, 1)\} \), let \( \mathbf{w} = (5, 8) \), and let \( \mathcal{C} \) be the standard basis. Plot these vectors with tails at the origin so that the vectors in \( \mathcal{B} \) are drawn in black, the vectors in \( \mathcal{C} \) are drawn in green, and \( \mathbf{w} \) is drawn in yellow. Find the \( \mathcal{B} \)-coordinates of all vectors in the drawing, and balance the \( \mathcal{B} \) coordinate estimates so that the equation

\[
\mathbf{w} = 5e_1 + 8e_2
\]

is valid in \( \mathcal{B} \)-coordinates (where \( e_1 \) and \( e_2 \) are the standard basis vectors). Finally, solve the problem algebraically and compare the exact answers to the estimates.

Our drawing appears in Figure 13, and it includes the parallelogram needed to estimate the \( \mathcal{B} \)-coordinates of \( \mathbf{w} \), which seems about 7 in the direction of \((1, 1)\) and somewhere around \(-2.5\) in the direction of \((1, 0)\), so our estimate is \((7, -2.5)_\mathcal{B}\). Since \( e_1 \) is the second element of \( \mathcal{B} \), its \( \mathcal{B} \)-coordinates are \((0, 1)_\mathcal{B}\). Drawing another parallelogram with a corner at \((0, 1)\) reveals the \( \mathcal{B} \)-coordinates of \( e_2 \) to be \((-1, 1)_\mathcal{B}\). Now we want the equation \( \mathbf{w} = 5e_1 + 8e_2 \) valid in \( \mathcal{B} \)-coordinates, and computing

\[
\begin{align*}
(7, -2.5)_\mathcal{B} &\neq 5(0, 1)_\mathcal{B} + 8(-1, 1)_\mathcal{B} \\
\end{align*}
\]

shows how good our estimates are. They are terrible, and probably they are signalling a mistake; we put the coordinates of \( e_2 \) in backwards. Since \((1, 1)\) is the first vector in \( \mathcal{B} \), the coordinate corresponding to this vector should be written first, and the multiple in the \((1, 0)\) direction should be written second, so the \( \mathcal{B} \)-coordinates of \( e_2 \) should be \((1, -1)_\mathcal{B}\). Trying the computation a second time

\[
\begin{align*}
(7, -2.5)_\mathcal{B} &\neq 5(0, 1)_\mathcal{B} + 8(1, -1)_\mathcal{B} \\
\end{align*}
\]

shows that we are getting close; if we tweak \((1, -1)_\mathcal{B}\) to \((\frac{7}{5}, -\frac{2}{5})_\mathcal{B}\) we are right on. So we submit the estimates \( \mathbf{w} \sim (7, -2.5)_\mathcal{B}, e_1 \sim (0, 1)_\mathcal{B}, \) and \( e_2 \sim (\frac{7}{5}, -\frac{2}{5})_\mathcal{B}. \)

Now we show how to solve the problem algebraically. Greenhorns are urged to compare the corresponding geometric solution in the previous paragraph. As above, we begin by seeking the \( \mathcal{B} \)-coordinates of \( \mathbf{w} \); writing the equation

\[
(5, 8) = \alpha(1, 1) + \beta(1, 0)
\]

is the algebraic equivalent of drawing the large parallelogram in Figure 13. Solving for \( \alpha \) and \( \beta \) is the algebraic equivalent of estimating what multiples of \((1, 1)\) and \((1, 0)\) give the edges of the parallelogram. Writing the equation above in terms of its first and second coordinates leads to the following system of equations:

\[
\begin{align*}
\alpha + \beta &= 5 \\
\alpha &= 8,
\end{align*}
\]

which gives \( \alpha = 8 \) and \( \beta = -3. \) The \( \mathcal{B} \)-coordinates of \((1, 0)\) are \((0, 1)_\mathcal{B}\), and we could also derive this algebraically by solving

\[
(1, 0) = \alpha(1, 1) + \beta(1, 0)
\]

for \( \alpha \) and \( \beta. \) Finally, the \( \mathcal{B} \)-coordinates of \((0, 1)\) are obtained by solving

\[
(0, 1) = \alpha(1, 1) + \beta(1, 0),
\]
giving $\alpha = 1$ and $\beta = -1$, so the $B$-coordinates of $(0,1)$ are $(1,-1)$. This algebraic solution gives us the exact answer, not an estimate, and we can now submit the exact solution $w = (8,-3)$, $e_1 = (0,1)$, and $e_2 = (1,-1)$. It never hurts to perform a check for error, and we can do so by verifying $w = 5e_1 + 8e_2$, so we calculate

$$(8,-3) = 5(0,1) + 8(1,-1)$$

and feel satisfied.

**Problem 11** In $R^2$, let $B = \{(-1,-1), (-1,0)\}$, let $w = (1,-1)$, and let $C = \{e_1, e_2\}$ be the standard basis. Plot these vectors with tails at the origin so that the vectors in $B$ are drawn in black, the vectors in $C$ are drawn in green, and $w$ is drawn in yellow. Find the $B$-coordinates of all vectors in the drawing, and balance the $B$ coordinate estimates so that the equation

$$w = e_1 - e_2$$

is valid in $B$-coordinates. Finally, solve the problem algebraically and compare the exact answers to the estimates.

**Problem 12** In $R^2$, let $B = \{(-1,0), (-1,1)\}$, let $w = (2,3)$, and let $C = \{e_1, e_2\}$ be the standard basis. Plot these vectors with tails at the origin so that the vectors in $B$ are drawn in black, the vectors in $C$ are drawn in green, and $w$ is drawn in yellow. Find the $B$-coordinates of all vectors in the drawing, and balance the $B$ coordinate estimates so that the equation

$$w = 2e_1 + 3e_2$$

is valid in $B$-coordinates. Solve the problem algebraically and compare the exact answers to the estimates.

### 2.4 Running with the algebra

Determining the $B$-coordinates of a vector algebraically is a direct appeal to the definition of coordinates; if $B = \{u, v\}$ and we know that $(\alpha, \beta)_B$ are the $B$-coordinates of $w$, then by the definition of coordinates, we have $w = \alpha u + \beta v$. Conversely, if we want to find the $B$-coordinates of $w$, then we write $w = \alpha u + \beta v$ and solve for $\alpha$ and $\beta$. There is nothing special about $R^2$ that makes definition work: as long as $B$ is any independent set, then every vector in its span determines unique scalars enabling us to write that vector as a linear combination of the vectors in $B$. If $B = \{(9,0,0),(3,5,0),(1,2,1)\}$, then every vector in $R^3$ acquires $B$-coordinates by writing the vector as $\alpha(9,0,0) + \beta(3,5,0) + \gamma(1,2,1)$; its $B$-coordinates are then $(\alpha, \beta, \gamma)_B$. If someone asks for the $B$-coordinates of $(0,4,5)$, we would write

$$(0,4,5) = \alpha(9,0,0) + \beta(3,5,0) + \gamma(1,2,1)$$

and solve for $\alpha$, $\beta$, and $\gamma$. If

$$B = \{(9,0,0,0),(3,5,0,0),(1,2,1,0),(8,8,8,8)\},$$

then every vector in $R^4$ acquires $B$-coordinates, and when asked for the $B$-coordinates of $(0,4,5,-1)$ we write

$$(0,4,5,-1) = \alpha(9,0,0,0) + \beta(3,5,0,0) + \gamma(1,2,1,0) + \epsilon(8,8,8,8)$$

and solve for $\alpha$, $\beta$, $\gamma$, and $\epsilon$. 


Example 5

Find the $B$-coordinates of $w = (2, -1)$ when $B = \{(1, 3), (-3, 1)\}$.

Write

\[(2, -1) = \alpha (1, 3) + \beta (-3, 1)\]

and solve for $\alpha$ and $\beta$. Solving the system of two equations in two unknowns reveals that $\alpha = -\frac{1}{10}$ and $\beta = -\frac{7}{10}$, so $(2, -1) = (-\frac{1}{10}, -\frac{7}{10})B$.

Example 6

In $C^2$, find the $B$-coordinates of $w = (i, -i)$ when $B = \{(i, 1), (-i, 1)\}$.

Write

\[(i, -i) = \alpha (i, 1) + \beta (-i, 1)\]

and solve for $\alpha$ and $\beta$. The first coordinates say that $i = \alpha i - \beta i$ and the second coordinates say that $-i = \alpha + \beta$; thus we set out to solve the system

\[
\begin{align*}
\alpha - \beta &= 1 \\
\alpha + \beta &= -i,
\end{align*}
\]

and we see the opportunity to add the equations, giving us $2\alpha = 1 - i$. Solving for $\alpha$ and substituting into the second equation gives

\[
\frac{1 - i}{2} + \beta = -i
\]

which says $\beta = -\frac{1+i}{2}$ and $(i, -i) = (\frac{1-i}{2}, -\frac{1+i}{2})B$.

Problem 13

Find the $B$-coordinates of $w = (1, -1)$ when $B = \{(1, 3), (-3, 1)\}$.

Problem 14

In $C^2$, find the $B$-coordinates of $w = (1, 1)$ when $B = \{(i, 1), (-i, 1)\}$.

Just as finding the coordinates of a vector in $R^2$ amounts to solving a system of two equations in two unknowns, finding the coordinates of a vector in a three dimensional space involves solving three equations for three unknowns. The reader is welcome to use whatever method they like for solving systems of equations; powerful calculators and computer algebra systems like Maple or Mathematica make this tedious task much more bearable.

Example 7

Find the $B$-coordinates of $w = (1, -1, 0)$ when $B = \{(1, 0, 1), (-2, 1, 1), (1, 1, -1)\}$.

Write

\[(1, -1, 0) = \alpha (1, 0, 1) + \beta (-2, 1, 1) + \gamma (1, 1, -1)\]

and solve for $\alpha$, $\beta$, and $\gamma$. From the three coordinates we get three equations

\[
\begin{align*}
\alpha - 2\beta + \gamma &= 1 \\
\beta + \gamma &= -1 \\
\alpha + \beta - \gamma &= 0
\end{align*}
\]

upon which we perform row operations to obtain the solution $\alpha = \frac{1}{5}$, $\beta = -\frac{3}{5}$, and $\gamma = -\frac{2}{5}$. It follow that $(1, -1, 0) = (\frac{1}{5}, -\frac{3}{5}, -\frac{2}{5})B$. 

Example 8

Find the $\mathcal{B}$-coordinates of $w = (i, -i, 1)$ when $\mathcal{B} = \{(1, i, 1), (-i, 1, 0), (0, 1, -1)\}$.

Write

$$(i, -i, 1) = \alpha(1, i, 1) + \beta(-i, 1, 0) + \gamma(0, 1, -1)$$

and solve for $\alpha$, $\beta$, and $\gamma$. This equation is equivalent to the system

$$\begin{align*}
\alpha - i\beta &= i \\
i\alpha + \beta + \gamma &= -i \\
\alpha - \gamma &= 1,
\end{align*}$$

which we proceed to row reduce to the form

$$\begin{align*}
i\alpha + \beta + \gamma &= -i \\
\beta + (1 + i)\gamma &= -2i \\
\gamma &= 1 - i.
\end{align*}$$

Substituting the value of $\gamma$ from the third line into the second, we can solve for $\beta$, after which we substitute both values of $\gamma$ and $\beta$ into the first line, obtaining $\alpha = 2 - i$, $\beta = -2 - 2i$, and $\gamma = 1 - i$. This is how we see that

$$(i, -i, 1) = (2 - i, -2 - 2i, 1 - i)_{\mathcal{B}}.$$

Problem 15

Find the $\mathcal{B}$-coordinates of $w = (1, 0, 1)$ when $\mathcal{B} = \{(1, 0, 1), (-2, 1, 1), (1, 1, -1)\}$.

Problem 16

Find the $\mathcal{B}$-coordinates of $w = (0, 2, 1)$ when $\mathcal{B} = \{(1, 0, 1), (-2, 1, 1), (1, 1, -1)\}$.

Problem 17

Find the $\mathcal{B}$-coordinates of $w = (1, 0, 1)$ when $\mathcal{B} = \{(1, i, 1), (-i, 1, 0), (0, 1, -1)\}$.

Problem 18

Find the $\mathcal{B}$-coordinates of $w = (-i, 1, 0)$ when $\mathcal{B} = \{(1, i, 1), (-i, 1, 0), (0, 1, -1)\}$.

Exercises

Pictures

Draw a picture depicting the following sets.

1. The set $A \equiv \{\alpha(-1, -1) + (1 - \alpha)(1, 1) : 0 \leq \alpha \leq 1\}$ as a subset of $R^2$.
2. The set $B \equiv \{(1, 2) + \alpha(1, 1) : \alpha \in R\}$ as a subset of $R^2$.
3. The set $C \equiv \{(1, 2) + \alpha(-1, 1) : \alpha \in R\}$ as a subset of $R^2$.
4. The set $D \equiv \{(1, 1) + \alpha(-1, 1) : \alpha \in R\}$ as a subset of $R^2$.
5. The set $E \equiv \{(1, 1) + \alpha(1, 1) : \alpha \in R\}$ as a subset of $R^2$.
6. The set $F \equiv \{(1, 0, 0) + \beta(0, 0, 1) : \alpha, \beta \in R\}$ as a subset of $R^3$.
7. The set $G \equiv \{(1, 1, 1) + \alpha(1, 0, 0) + \beta(0, 0, 1) : \alpha, \beta \in R\}$ as a subset of $R^3$.
8. The set $H \equiv \{(1, 1, 1) + \alpha(1, 0, 0) + \beta(0, 0, 1) : \alpha, \beta \in R\}$ as a subset of $R^3$.
9. The set $I \equiv \{(1, 1, 1) + \beta(0, 1, 1) : \alpha, \beta \in R\}$ as a subset of $R^3$.
10. $B \cap C$
11. $A \cap C$
12. $A \cap D$
13. $F \cap G$
14. $F \cap I$

Equivalence relations

The set $R^2$ can be partitioned into the lines parallel to the $x$-axis. Let $Q$ denote the set of consisting of these lines, so we could reiterate the definition of $Q$ by writing

$$Q = \{L \subset R^2 : L \text{ is a line with slope } 0\}.$$

If $u \in R^2$, then $u$ is on exactly one of these lines; lets let $L_u$ denote the line that contains $u$.

15. Find the condition on the coordinates of $u$ and $v$ that is equivalent to $L_u = L_v$.

16. Subsets of $R^2$ can be added to obtain a new subset; to add two sets $A$ and $B$, form the set of all possible sums where one summand is in $A$ and the other is in $B$, i.e. form the set

$$\{a + b : a \in A, b \in B\}.$$
Chapter 2 • Coordinates

Show that $L_u + L_v = L_{u+v}$.
17. Assume that $\mathcal{L}$ denotes the set of all lines in $R^2$. Find two elements of $\mathcal{L}$ for which their sum is not an element of $\mathcal{L}$; i.e. find two lines $L_1$ and $L_2$ such that $L_1 + L_2$ is not a line.
18. Let $Q$ now denote a partition of $R^2$ into lines (a generalization of the previous $Q$); so $Q$ is a set consisting of disjoint lines so that every vector in $R^2$ finds itself on one of these lines. Show that if $L_1$ and $L_2$ are in $Q$, then so is their sum.

Examples

19. an invertible function $f : R^2 \to R^2$ so that $f - f^{-1} = 0$.
20. an invertible function $f : R^2 \to R^2$ so that $f + f^{-1} = 0$.
21. a function $f : R^2 \to R^2$ so that $f \circ f = f$.
22. a function $f : R^2 \to R^2$ with no invariant subspaces other than $R^2$.
23. a basis $\mathcal{B}$ where the $\mathcal{B}$-coordinates of $(2, 5)$ are $(0, 1)_B$.
24. a basis $\mathcal{B}$ where the $\mathcal{B}$-coordinates of $(2, 5)$ are $(1, 1)_B$.
25. two disjoint bases $\mathcal{B}$ and $\mathcal{C}$ where the $\mathcal{C}$-coordinates of $(1, 1)_B$ are $(1, 1)_C$.
26. a basis $\mathcal{B}$ of $R^3$ whose first two vectors span the $yz$-plane.
27. a basis $\mathcal{B}$ of $R^3$ whose first two vectors span the plane $\{(x, y, z) : 2x + 3y - z = 0\}$.

Counterexamples

28. If $\mathcal{B}$ is a nonstandard basis of $R^2$ and $u \in R^2$, then the $\mathcal{B}$-coordinates of $u$ differ from its standard coordinates.
29. If $u \in R^2$ is a vector with length less than 1 and $\mathcal{B}$ is a basis of $R^2$, then the $\mathcal{B}$-coordinates of $u$ are both less than 1.
30. Every function $f : R^2 \to R^2$ can be expresses with a formula of the form

$$f(x, y)_B = (ax, by)_B$$

relative to some basis $\mathcal{B}$ and some scalars $a, b \in R$.

Et cetera

When we are dealing with a space of vectors and using only real scalars, we call the space a real vector space. When we allow the scalars to be complex numbers, the space of vectors is called a complex vector space. The spaces $C^n$ are complex vector spaces, but since $R \subset C$ they are also real vector spaces. Unless specifically instructed to think otherwise, you should consider $C^n$ as a complex vector space and use complex scalars when working in this space.

31. The space $C^2$ is a complex vector space; find its dimension and find a basis.
32. The space $C^2$ is also a real vector space; find its dimension as a real vector space (it’s not 2), and find a basis of $C^2$ as a real vector space.

Find the $\mathcal{B}$-coordinates for $w$ when...
33. $\mathcal{B} = \{(0, 1), (1, 0)\}$ and $w = (2, 1)$
34. $\mathcal{B} = \{(0, 1), (1, 0)\}$ and $w = (a, b)$
35. $\mathcal{B} = \{(1, 1), (1, -1)\}$ and $w = (0, 1)$
36. $\mathcal{B} = \{(1, -1), (1, 1)\}$ and $w = (1, 0)$
37. $\mathcal{B} = \{(0, 1, 0), (0, 0, 1), (1, 0, 0)\}$ and $w = (2, 3, 4)$
38. $\mathcal{B} = \{(0, 1, 0), (0, 0, 1), (1, 0, 0)\}$ and $w = (a, b, c)$
39. $\mathcal{B} = \{(0, 0, 1), (0, 1, 0), (1, 0, 0)\}$ and $w = (a, b, c)$
40. $\mathcal{B} = \{(0, i), (i, 1)\}$ and $w = (1, 0)$
41. $\mathcal{B} = \{(0, i), (i, 1), (1, i + 1)\}$ and $w = (1, 0)$ in $C^2$ considered as a real vector space.
In the following exercises, introduce a basis $\mathcal{B}$, if possible, so that the formula for the vector field is of the form

$$f(x, y)_{\mathcal{B}} = (ax, by)_{\mathcal{B}},$$

for some scalars $a, b \in \mathbb{R}$. If you don’t think such a basis exists, say why.

42. 43. 44. 45.

46. 47. 48. 49.

More Exercises
Chapter 3
Matrices

3.1 The leading role

We have already met the star of our subject, but let us now make a formal introduction. Whenever mathematical objects arise it is not only the objects themselves that merit attention, but also the mappings between these objects, and in particular, the mappings that preserve the structure inherent in the objects. The objects in the study of linear algebra are the spaces like $R^2$, $C^2$, $R^3$, $C^3$, on up to $R^n$, $C^n$, and the structure found in these objects is vector addition and scalar multiplication. The mappings that preserve this structure play the leading role in linear algebra, and these are the linear transformations. What is meant by preserving structure? Using algebra we can express this very succinctly; a function $T$ is linear if and only if

$$T(\alpha u + \beta v) = \alpha T(u) + \beta T(v)$$

for every linear combination in the domain. This equality says that a linear combination of $u$ and $v$ in the domain of $T$ should be taken to the exact corresponding linear combination of $T(u)$ and $T(v)$ in the codomain. We now begin to denote linear transformations with capitall letters like $T$, and except in places where it is specifically said otherwise, all the functions that follow will be linear.

The linear transformations that also happen to be invertible functions are called isomorphisms, and these isomorphisms provide us with a way to identify two spaces that are perhaps very different sets, but are structurally identical. When an isomorphism exists between two spaces we say that the spaces are isomorphic. For example, any two planes in $R^3$ that contain the origin, while they need not be the same plane, they are isomorphic and they are both isomorphic to $R^2$.

Problem 1

The definition of a linear transformation implies that $T(o + o) = T(o) + T(o)$. Use this to show that, if $T$ is linear, then $T(o) = o$.

The jist of problem 1 is that the graph of a linear transformation must pass through the origin. The term 'linear' comes from the shape of the graph; if $(u, T(u))$ and $(v, T(v))$ are two elements of the graph of $T$, then the equality

$$\alpha(u, T(u)) + \beta(u, T(u)) = (\alpha u + \beta v, T(\alpha u + \beta v))$$

shows that the graph is a subspace. Thus the graphs of linear transformations in $R^2$ are lines containing the origin, and graphs of linear transformations in $R^3$ are either planes or lines containing the origin. (The origin all by itself is also the graph of a linear transformation, but not a very interesting one.)

Problem 2

Give an example of a function $f : R \rightarrow R$ for which $f(0) = 0$ but $f$ is not a linear transformation.

Two points determine a line, so if we know that the graph of some function is a line and we know two points on the graph, then that function is completely determined. If we are given a function
3.2 • The matrix of a transformation

\( T : R \to R \) that is linear (a shorthand way of saying \( T \) is a linear transformation), then we know that its graph is a line and we know \((0,0)\) is on that line. It follows that a linear transformation \( T : R \to R \) is completely determined as soon as we know just one non-zero point on its graph.

**Problem 3**
Assume \( T : R \to R \) is a linear transformation such that \( T(3) = 2 \). Draw the graph of \( T \), then write a formula for \( T \).

Three points, not all on the same line, determine a plane. If we know that a particular plane is the graph of a linear transformation, then we know that the vector \( \mathbf{0} \) is in that plane. Thus if we know two other points in the plane that are not multiples of each other, we can completely determine the linear function. What we are describing geometrically has a corresponding algebraic formulation. If \( \mathbf{u} \) and \( \mathbf{v} \) form a basis of \( R^2 \), then any linear transformation whose domain is \( R^2 \) is completely determined by what it does to just these two vectors. How can knowing what a transformation \( T \) does to the two vectors \( \mathbf{u} \) and \( \mathbf{v} \) have anything to do with what it does to a generic vector \( \mathbf{w} \)? The answer has two parts; first, every vector \( \mathbf{w} \) can be written as a linear combination of \( \mathbf{u} \) and \( \mathbf{v} \), and second, a linear transformation must take a linear combination of \( \mathbf{u} \) and \( \mathbf{v} \) to the exact corresponding combination of \( T(\mathbf{u}) \) and \( T(\mathbf{v}) \). Thus if we know only what \( T(\mathbf{u}) \) and \( T(\mathbf{v}) \) are, then we can compute \( \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) \), which, by the linearity of \( T \), is the same as \( T(\alpha \mathbf{u} + \beta \mathbf{v}) \).

**Example 1**
Assume \( T : R^2 \to R \) is linear, \( T(1,0) = 2 \) and \( T(1,1) = 1 \). Find \( T(5,1) \).

Let the basis consist of the two vectors for which the action of \( T \) is known; i.e. let \( \mathcal{B} = \{(1,0),(1,1)\} \). Now we need to find the \( \mathcal{B} \)-coordinates of \((5,1)\), so write

\[
(5,1) = \alpha(1,0) + \beta(1,1)
\]

and solve for \( \alpha \) and \( \beta \), getting \( \alpha = 4 \) and \( \beta = 1 \). Put these values into the equation above to get

\[
(5,1) = 4(1,0) + 1(1,1),
\]

and apply \( T \) to both sides

\[
T(5,1) = T(4(1,0) + 1(1,1)) = 4T(1,0) + 1T(1,1).
\]

Substituting the given values for \( T(1,0) \) and \( T(1,1) \) into the equation above gives \( T(5,1) = 9 \).

**Problem 4**
Assume \( T : R^2 \to R \) is linear, \( T(1,-1) = 3 \) and \( T(1,1) = 1 \). Find \( T(2,-1) \).

**Example 2**
Assume \( T : R^2 \to R^2 \) is linear, \( T(1,2) = (3,1) \) and \( T(-1,1) = (4,1) \). Find \( T(1,0) \).

Let \( \mathcal{B} = \{(1,2),(-1,1)\} \) and get the \( \mathcal{B} \)-coordinates of \((1,0)\); so write

\[
(1,0) = \alpha(1,2) + \beta(-1,1)
\]

and solve for \( \alpha \) and \( \beta \), getting \( \alpha = \frac{1}{3} \) and \( \beta = -\frac{2}{3} \). Substituting in for \( \alpha \) and \( \beta \) gives

\[
(1,0) = \frac{1}{3}(1,2) - \frac{2}{3}(-1,1),
\]

and applying \( T \) to both sides yields the answer

\[
T(1,0) = \frac{1}{3}T(1,2) - \frac{2}{3}T(-1,1) = \frac{1}{3}(3,1) - \frac{2}{3}(4,1) = (-\frac{5}{3}, -\frac{1}{3}).
\]
Problem 5 Assume $T : R^2 \rightarrow R^2$ is linear, $T(2,0) = (1,2)$ and $T(8,1) = (0,0)$. Find $T(1,0)$.

3.2 The matrix of a transformation

When a function $T : R^2 \rightarrow R^2$ is linear, the coordinates of the output $T(x,y)_B$ always assume the form $ax + by$, where $a$ and $b$ are scalars that are determined by the transformation. Since a transformation such as $T$ will have two coordinates in the output, we get a total of four scalars that determine the formula for $T$; the formula looks like $T(x,y)_B = (ax + by, cx + dy)_B$. There is an established convention for encoding this formula in terms of matrices; we first write the elements of $R^2$ vertically, instead of horizontally, and we then proclaim the definition

$$
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y
\end{pmatrix}_{B} \equiv \begin{pmatrix}
a & c \\
b & d
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y
\end{pmatrix}_{B} + \begin{pmatrix}
a & b \\
c & d
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y
\end{pmatrix}_{B}.
\end{equation}

Notice that the right side of this equation is the linear combination $(ax + by, cx + dy)_B$ (written vertically). We encourage you to think of the left side of the equation as $T(x,y)_B$, with the $2 \times 2$ matrix being $T$. We would also like to suggest a way of remembering how to multiply the matrices; we are taking a linear combination of the columns of the leftmost matrix. If the matrix has three columns, instead of two, then we would need a vector in $R^3$ to multiply on the right, like

$$
\begin{pmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B} \equiv \begin{pmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B} + \begin{pmatrix}
a & b & c \\
d & e & f \\
g & h & i
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B}.
\end{equation}

or

$$
\begin{pmatrix}
a & b & c \\
d & e & f
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B} \equiv \begin{pmatrix}
a & b & c \\
d & e & f
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B} + \begin{pmatrix}
a & b & c \\
d & e & f
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y \\
z
\end{pmatrix}_{B}.
\end{equation}

The multiplication

$$
\begin{pmatrix}
a & b & c \\
d & e & f
\end{pmatrix}_{B} \begin{pmatrix}
x \\
y
\end{pmatrix}_{B}
$$

is not defined. As long as the number of columns is the same as the number of coordinates in the column vector, the definition is valid, and the general form of the definition looks like

$$
(a_{ij})_B(x_j)_B \equiv \sum_{j=1}^{m} x_j(a_{ij})_B.
$$

If we have a transformation $T : R^2 \rightarrow R^2$, the entries in the matrix

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

are determined by what $T$ does to two vectors; the first column contains the $B$-coordinates of $T(u)$ and the second column contains the $B$-coordinates of $T(v)$, where $B = \{u, v\}$. We refer to this matrix as the $B$-matrix for $T$. When a matrix is given with no subscript, it is considered to be a $B$-matrix with $B$ being the standard basis, and we call it the standard matrix for $T$.

Example 3 Assume $T : R^2 \rightarrow R^2$ is the transformation that takes $u$ to $\frac{u}{3}$ and takes $v$ to $\frac{v}{2}$ in figure 1. With $B = \{u, v\}$, find the $B$-matrix for $T$. Use this matrix to give a formula for $T$. 

---

Fig. 1:
To get the $B$-matrix for $T$, we write the $B$-coordinates of $T(u)$ in the first column, and the $B$-coordinates of $T(v)$ in the second column. Since $T(u) = \frac{5}{4}$, we write

$$\frac{u}{3} = \alpha u + \beta v$$

and solve for $\alpha$ and $\beta$, getting $\alpha = \frac{1}{3}$ and $\beta = 0$. This gives us the first column of the solution, and to get the second column we find the $B$-coordinates of $T(v)$, i.e. of the vector $\frac{1}{2}$, which are 0 and $\frac{1}{2}$, so the $B$-matrix of $T$ is

$$\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}_B.$$

The matrix form of the formula for $T$ then becomes

$$\begin{pmatrix} \frac{1}{3} & 0 \\ 0 & \frac{1}{2} \end{pmatrix}_B \begin{pmatrix} x \\ y \end{pmatrix}_B = x \begin{pmatrix} \frac{1}{3} \\ 0 \end{pmatrix}_B + y \begin{pmatrix} 0 \\ \frac{1}{2} \end{pmatrix}_B = \begin{pmatrix} \frac{1}{3}x \\ \frac{1}{2}y \end{pmatrix}_B,$$

which corresponds to the formula $T(x, y)_B = (\frac{1}{3}x, \frac{1}{2}y)_B$.

**Example 4**

Assume that $T$ is the same transformation described in example 3, and let

$$C = \{(\frac{1}{2}, -\frac{1}{2})_B, (\frac{1}{2}, \frac{1}{2})_B\}.$$

Find the $C$-matrix for $T$, and give the corresponding formula for $T$.

We apply the transformation to the vectors in $C$ and write the resulting vectors in the columns of the matrix in terms of their $C$-coordinates. Applying $T$ to the first vector gives

$$T(\frac{1}{2}, -\frac{1}{2})_B = (\frac{1}{3}, \frac{1}{2}, -\frac{1}{2})_B = (\frac{1}{6}, -\frac{1}{4})_B,$$

and this vector goes in the first column in terms of its $C$-coordinates. To find its $C$-coordinates, write

$$\begin{pmatrix} \frac{1}{6} \\ -\frac{1}{4} \end{pmatrix}_B = \alpha (\frac{1}{2}, -\frac{1}{2})_B + \beta (\frac{1}{2}, \frac{1}{2})_B,$$

and solve for $\alpha$ and $\beta$. This equation leads to the system

$$\begin{align*}
\alpha + \beta &= \frac{1}{6} \\
-\alpha + \beta &= -\frac{1}{4}.
\end{align*}$$

which has the solution $\alpha = \frac{5}{12}$ and $\beta = \frac{1}{12}$.

Performing the same process on the second vector, we first compute

$$T(\frac{1}{2}, \frac{1}{2})_B = (\frac{1}{3}, \frac{1}{2}, \frac{1}{2})_B = (\frac{1}{6}, \frac{1}{4})_B,$$

then convert the result into $C$-coordinates; write

$$\begin{pmatrix} \frac{1}{6} \\ \frac{1}{4} \end{pmatrix}_B = \alpha (\frac{1}{2}, -\frac{1}{2})_B + \beta (\frac{1}{2}, \frac{1}{2})_B.$$
and solve for $\alpha$ and $\beta$. We get $\alpha = -\frac{1}{12}$ and $\beta = \frac{5}{12}$. Thus we have that the $C$-matrix of $T$ is

$$\begin{pmatrix}
\frac{5}{12} & -\frac{1}{12} \\
-\frac{1}{12} & \frac{5}{12}
\end{pmatrix}_C.$$

The matrix form of the formula for $T$ is thus

$$\begin{pmatrix}
\frac{5}{12} & -\frac{1}{12} \\
-\frac{1}{12} & \frac{5}{12}
\end{pmatrix}_C \begin{pmatrix}x \\ y\end{pmatrix}_C = x \begin{pmatrix} \frac{5}{12} \\ -\frac{1}{12} \end{pmatrix}_C + y \begin{pmatrix} -\frac{1}{12} \\ \frac{5}{12} \end{pmatrix}_C = \frac{5x}{12} - \frac{y}{12} + \frac{5y}{12},$$

and the corresponding horizontal formula is $T(x, y)_C = (\frac{5x}{12} - \frac{y}{12} + \frac{5y}{12})_C$. The function in examples 3 and 4 was taken from example 1 on page 16, where the same problem was worked geometrically.

The reader is urged to review that material in order to completely understand what was just done. Indeed, the rest of this book depends on absorbing the material in this section, so work hard now and reap the benefits later.

Problem 6

Assume $T$ is the transformation that, in figure 2, takes $u$ to $\frac{u}{2}$ and takes $v$ to itself. Find the $B$-matrix of $T$ when $B = \{u, v\}$, and find the $C$-matrix of $T$ when $C = \{(\frac{1}{2}, \frac{1}{2})_B, (\frac{1}{2}, \frac{1}{2})_B\}$. Give the formula for $T$ relative to both coordinate systems.

Problem 7

Assume $T$ is the transformation that, in figure 2, takes $u$ to $v$ and takes $v$ to $u$. Find the $B$-matrix of $T$ when $B = \{u, v\}$, and find the $C$-matrix of $T$ when $C = \{(\frac{1}{2}, \frac{1}{2})_B, (\frac{1}{2}, \frac{1}{2})_B\}$. Give the formula for $T$ relative to both coordinate systems.

Here is a tip on how one checks their answers to questions like those above. We arrived at the answer of

$$\begin{pmatrix}
\frac{5}{12} & -\frac{1}{12} \\
-\frac{1}{12} & \frac{5}{12}
\end{pmatrix}_C$$

for the $C$-matrix of a transformation $T$ that maps the vector $v$ to $\frac{v}{2}$; we can find the $C$-coordinates $(\alpha, \beta)_C$ of $v$ and verify that

$$\begin{pmatrix}
\frac{5}{12} & -\frac{1}{12} \\
-\frac{1}{12} & \frac{5}{12}
\end{pmatrix}_C \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_C = \frac{1}{2} \begin{pmatrix} \alpha \\ \beta \end{pmatrix}_C.$$

To get these $C$-coordinates, we write $v = \alpha(\frac{1}{2}, \frac{1}{2})_B + \beta(\frac{1}{2}, \frac{1}{2})_B$, i.e. we write

$$ (0, 1)_B = \alpha \left( \frac{1}{2}, \frac{1}{2} \right)_B + \beta \left( \frac{1}{2}, \frac{1}{2} \right)_B,$$

and solve for $\alpha$ and $\beta$, getting $\alpha = -1$ and $\beta = 1$. We then compute

$$\begin{pmatrix}
\frac{5}{12} & -\frac{1}{12} \\
-\frac{1}{12} & \frac{5}{12}
\end{pmatrix}_C \begin{pmatrix} -1 \\ 1 \end{pmatrix}_C = \begin{pmatrix} -\frac{1}{2} \\ \frac{1}{2} \end{pmatrix}_C,$$

and are satisfied.

Problem 8

Demonstrate that your answers in problems 6 and 7 are correct by performing the check outlined in the paragraph above.

3.3 Matrix algebra

There are various algebraic operations that can be performed on linear transformations. Two transformations $S$ and $T$ with the same domain and codomain can be added to obtain a transformation
$S + T$, any transformation can be multiplied by a scalar, and, as long as the range of $T$ is inside the
domain of $S$, two transformations $S$ and $T$ can be composed to obtain $S \circ T$, which is thought of as
a type of multiplication. These three operations are defined by the formulæ

$$(S + T)(u) = S(u) + T(u) \quad (\alpha T)(u) = \alpha T(u) \quad (S \circ T)(u) = S(T(u)).$$

For example, if $S, T : \mathbb{R}^2 \to \mathbb{R}^2$ are defined by

$$S(x, y) = (2x + y, x - 4y), \quad T(x, y) = (7y, 5x - 3y),$$

and we have the scalar $\alpha = -1$ then

$$(S + T)(x, y) = (2x + 8y, 6x - 7y), \quad (\alpha T)(x, y) = (-7y, -5x + 3y),$$

and

$$(S \circ T)(x, y) = S(7y, 5x - 3y) = (2(7y) + (5x - 3y), 7y - 4(5x - 3y)).$$

**Problem 9** Assume that $S$ and $T$ are given by the equations in the paragraph above, and $\mathcal{B} = \{(1, 0), (0, 1)\}$ is the standard basis. Find the $\mathcal{B}$-matrices of $S$, $T$, and $S + T$.

**Problem 10** Assume that $S$ and $T$ are given by the equations in the paragraph above; give the formula for $T \circ S$ and show that it is different than $S \circ T$.

**Example 5** Assume that $S, T : \mathbb{R}^2 \to \mathbb{R}^2$ are arbitrary linear transformations and $\mathcal{B} = \{\mathbf{u}, \mathbf{v}\}$ is an arbitrary basis of $\mathbb{R}^2$. Find how to obtain the $\mathcal{B}$-matrix of $S \circ T$ in terms of the $\mathcal{B}$-matrices of $S$ and $T$.

Assume the $\mathcal{B}$-matrices of $S$ and $T$ are

$$
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}_\mathcal{B} \quad \text{and} \quad 
\begin{pmatrix}
e & f \\
g & h
\end{pmatrix}_\mathcal{B},
$$

respectively. The definition of the $\mathcal{B}$-matrix of $S \circ T$ tells us to compute $S(T(\mathbf{u}))$ and $S(T(\mathbf{v}))$, and put these in the columns of the matrix in terms of their $\mathcal{B}$-coordinates. The $\mathcal{B}$-coordinates of $T(\mathbf{u})$ and $T(\mathbf{v})$ appear in the rightmost matrix above, so we can compute $S(T(\mathbf{u}))$ and $S(T(\mathbf{v}))$ by performing the matrix products

$$
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}_\mathcal{B} 
\begin{pmatrix}
e & f \\
g & h
\end{pmatrix}_\mathcal{B} \quad \text{and} \quad 
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix}_\mathcal{B} 
\begin{pmatrix}
e & f \\
g & h
\end{pmatrix}_\mathcal{B}.
$$

Putting the results of these products in the first and second columns gives us the matrix

$$
\begin{pmatrix}
ac + bg & af + bh \\
ce + dg & cf + dh
\end{pmatrix}_\mathcal{B},
$$

which the reader may recognize as the usual product of matrices

$$
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix} 
\begin{pmatrix}
e & f \\
g & h
\end{pmatrix} =
\begin{pmatrix}
ac + bg & af + bh \\
ce + dg & cf + dh
\end{pmatrix}.
$$

This is no coincidence; matrix multiplication is defined so that it will correspond to the composition of the associated linear transformations. A good way to remember how to multiply matrices is to think of applying the matrix on the left to each column of the matrix on the right.
Problem 11 Assume that $S, T : R^2 \rightarrow R^2$ are arbitrary linear transformations and $\mathcal{B} = \{u, v\}$ is an arbitrary basis of $R^2$. Find how to obtain the $\mathcal{B}$-matrix of $S + T$ in terms of the $\mathcal{B}$-matrices of $S$ and $T$, and justify your answer.

Problem 12 Assume that $T : R^2 \rightarrow R^2$ is an arbitrary linear transformation, $\alpha$ is a scalar, and $\mathcal{B} = \{u, v\}$ is a basis of $R^2$. Find how to obtain the $\mathcal{B}$-matrix of $\alpha T$ in terms of the $\mathcal{B}$-matrix of $T$, and justify your answer.

When multiplying two matrices, there is no restriction on the number of columns in the right most matrix, but those column vectors need to have the same number of coordinates as the number of columns in the left most matrix. In other words, the number of columns in the left most matrix must equal the number of rows in the right most matrix. When this is the case, then the general equation for multiplying two matrices takes the form

$$(a_{ij})(b_{ij}) = \sum_{k=1}^{m} a_{ik}b_{kj},$$

which is nothing more than a concise way of saying ‘apply $a_{ij}$ to the columns of $b_{ij}$ to get the columns of the product’.

Problem 13 Assume that $S, T : R^3 \rightarrow R^3$ are arbitrary linear transformations and $\mathcal{B} = \{u, v, w\}$ is an arbitrary basis of $R^3$. Find how to obtain the $\mathcal{B}$-matrix of $S \circ T$ in terms of the $\mathcal{B}$-matrices of $S$ and $T$.

3.4 Some matrix forms

The advantage of associating a transformation with a matrix comes from being able to read off information about the transformation from the form that its matrix assumes. The $\mathcal{B}$-matrix of a transformation will change as the basis $\mathcal{B}$ changes, and the strategy is to select a basis that reveals as much as possible about the transformation.

The diagonal from the northwest to southeast corner is called the main diagonal of a matrix, and when the matrix consists of zeros off of this diagonal, we call it a diagonal matrix. The following are all diagonal matrices;

$$
\begin{pmatrix}
6 & 0 \\
0 & -8
\end{pmatrix}_B, \begin{pmatrix}
0 & 0 & 0 \\
0 & 8 & 0 \\
0 & 0 & 3
\end{pmatrix}_B, \begin{pmatrix}
5 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}_B, \begin{pmatrix}
2 & 0 & 0 & 0 \\
0 & 3 & 0 & 0 \\
0 & 0 & 4 & 0 \\
0 & 0 & 0 & 5
\end{pmatrix}_B.
$$

When it is possible to find a basis $\mathcal{B}$ so that a transformation’s $\mathcal{B}$-matrix is diagonal, we say that the transformation is diagonalizable. Not every transformation is diagonalizable, but when one is, this is the most desirable matrix form to obtain because it reveals all of the eigenvalues and all of the eigenvectors. The eigenvalues appear on the main diagonal, and the vectors in the basis are the eigenvectors. The invertibility of a diagonalizable transformation is also immediately apparent from a glance at its diagonal form; the transformation is invertible exactly when the scalar 0 is not an eigenvalue. Of the diagonal matrices presented above, two are invertible, the first and the last. The middle two fail to be invertible because 0 appears on the main diagonal, and hence 0 is one of the eigenvalues. A matrix (or a transformation) is called singular when it is not invertible.

Problem 14 Assume $\mathcal{B} = \{(1,2,3),(1,-2,0),(-1,0,0)\}$ is a basis of $R^3$, and $T : R^3 \rightarrow R^3$ is a transformation with $\mathcal{B}$-matrix

$$
\begin{pmatrix}
0 & 0 & 0 \\
0 & 8 & 0 \\
0 & 0 & 3
\end{pmatrix}_B.
$$
Find the standard coordinates of two distinct vectors that are both taken to \( \mathbf{0} \). Conclude that the transformation is not one-to-one.

If \( T \) and \( S \) are two transformations whose \( \mathcal{B} \) matrices are known, then it is possible to find the \( \mathcal{B} \) matrix of \( T^4 + T^2S + ST - S^8 \) by performing the corresponding matrix algebra discussed in the previous section. If the \( \mathcal{B} \)-matrices of \( S \) and \( T \) are not in any particular form, then the computation of the matrix entries for this algebraic expression will be extremely messy. The greatest advantage of the diagonal form is that performing algebraic expressions becomes easy; if the \( \mathcal{B} \)-matrices of \( S \) and \( T \) are both in diagonal form, then the \( \mathcal{B} \) matrix of an algebraic expression in \( S \) and \( T \) will also be in diagonal form, and the diagonal entries can be obtained by performing the same algebraic expression on the diagonal entries of \( S \) and \( T \). For example, if \( s \) is the northwest entry of the diagonal matrix \( S \), and \( t \) is the northwest entry of the diagonal matrix \( T \), then the northwest entry of \( T^4 + T^2S + ST - S^8 \) is \( t^4 + t^2s + st - s^8 \).

**Problem 15** Assume the \( \mathcal{B} \)-matrices of \( S \) and \( T \) are

\[
\begin{pmatrix}
  0 & 0 & 0 \\
  0 & 8 & 0 \\
  0 & 0 & 3
\end{pmatrix}_B \quad \text{and} \quad \begin{pmatrix}
  0 & 0 & 0 \\
  0 & 1 & 0 \\
  0 & 0 & -1
\end{pmatrix}_B,
\]
respectively. Find all the eigenvalues of \( T^4 + T^2S + ST \).

Not every transformation is diagonalizable, and it is even rarer for two transformations to be diagonalizable relative to the same basis; when this is possible, we say that the two transformations are **simultaneously diagonalizable**. It could happen that both \( S \) and \( T \) are diagonalizable, but if they are diagonalized relative to different bases, there will not be a simple way of computing algebraic expressions like \( T^4 + T^2S + ST - S^8 \). There is another matrix form that comes to the rescue in this case; if a matrix has all zero entries below the main diagonal, we say that the matrix is **upper triangular**. The following matrices are all upper triangular:

\[
\begin{pmatrix}
  6 & 9 \\
  0 & -8
\end{pmatrix}_B \quad \begin{pmatrix}
  1 & 2 & 3 \\
  0 & 4 & 5 \\
  0 & 0 & 6
\end{pmatrix}_B \quad \begin{pmatrix}
  0 & 0 & 0 \\
  0 & 0 & 0 \\
  0 & 0 & 0
\end{pmatrix}_B \quad \begin{pmatrix}
  2 & 0 & 0 \\
  0 & 3 & 0 \\
  0 & 0 & 4
\end{pmatrix}_B \quad \begin{pmatrix}
  0 & 0 & 0 \\
  0 & 0 & 5 \\
  0 & 0 & 0
\end{pmatrix}_B.
\]

The two rightmost examples illustrate that diagonal matrices are upper triangular, but not every upper triangular matrix is diagonal. When a basis exists where the corresponding matrix of a transformation is triangular, we say that the transformation is **triangularizable**. If a single basis exists where the corresponding matrices of two or more transformations are triangular, we say that those transformations are **simultaneously triangularizable**.

When a transformation has a \( \mathcal{B} \)-matrix in upper triangular form, all the eigenvalues are visible on the main diagonal, just like they are for diagonal matrices. Moreover, even though computing the entries above the main diagonal of an algebraic expression is a mess, computing the diagonal entries works just like it does for diagonal matrices.

**Lower triangular** matrices are defined in analogy with upper triangular matrices. All of the computational advantages enjoyed by upper triangular matrices are shared by their lower triangular counterparts. In particular, the eigenvalues are visible on the main diagonal of a lower triangular matrix.

**Problem 16** Assume the \( \mathcal{B} \)-matrices of \( S \) and \( T \) are

\[
\begin{pmatrix}
  1 & 2 & 3 \\
  0 & 4 & 5 \\
  0 & 0 & 6
\end{pmatrix}_B \quad \text{and} \quad \begin{pmatrix}
  -1 & 2 & 3 \\
  0 & -4 & 5 \\
  0 & 0 & -6
\end{pmatrix}_B,
\]
respectively. Find the diagonal entries and the northeast entry of the \( \mathcal{B} \)-matrix for \( T^3 + S^3 \).
3.5 Rank and nullity

A linear transformation comes with a domain, codomain and a range, just like any other function. For linear functions there is an important subspace of the domain, called the nullspace (or kernel) of the linear transformation, which is the set of vectors that are taken to the zero vector by the transformation. The nullspace of \( T : \mathcal{V} \rightarrow \mathcal{W} \) is exactly the set

\[ \{ v \in \mathcal{V} : T(v) = 0 \} . \]

**Problem 17** Show that the nullspace and the range of a transformation are subspaces. (To show that the nullspace is a subspace, show that linear combinations of elements in the nullspace remain inside the nullspace.)

As subspaces, the nullspace and the range of a transformation have dimensions. The dimension of the nullspace is called the nullity of the transformation, and the dimension of the range is called the rank of the transformation. There is a relationship between these numbers that is the content of the rank and nullity theorem: the sum of the rank and nullity always equals the dimension of the domain. Any matrix form for a particular transformation reveals, to an extent, the range of the transformation; the span of the column vectors is the range. What may be somewhat concealed is the rank of the transformation, because it may not be clear if the column vectors are independent, and if they are not, what the dependency is. However, if the matrix of the transformation is in diagonal form, nothing is concealed (do problem 18).

**Example 6** Consider the transformation with standard matrix

\[
\begin{pmatrix}
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
\end{pmatrix}.
\]

Find the range and the nullspace of the transformation.

The range of the transformation is the span of the vectors in the columns, which is the one dimensional subspace spanned by \((1, 1, 1, 1)\). Now we know that the rank is one, and the rank and nullity theorem says that the rank and nullity add up to four, so the nullity is three. We will know the nullspace if we can find three independent vectors \((x, y, z, w)\) for which

\[
\begin{pmatrix}
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
1 & -1 & 1 & -1 \\
\end{pmatrix} \begin{pmatrix}
x \\
y \\
z \\
w \\
\end{pmatrix} = \begin{pmatrix}
0 \\
0 \\
0 \\
0 \\
\end{pmatrix}.
\]

The dependencies amongst the columns suggests several solutions; since the second column is the negative of the first column, \((1, 1, 0, 0)\) is a solution. Since the third column equals the first, \((1, 0, -1, 0)\) is a solution. The dependency between the first and last columns gives us the solution \((1, 0, 0, 1)\). Since none of these vectors is in the plane determined by the other two, the vectors are independent, and the nullspace is their span.

**Problem 18** Consider the transformations whose standard matrices are given below. Find the range and nullspace of each transformation, then verify the rank and nullity theorem.

\[
i) \begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 6 \\
\end{pmatrix} \quad ii) \begin{pmatrix}
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
0 & 0 & 3 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix} \quad iii) \begin{pmatrix}
0 & -1 & 1 & -1 \\
0 & 0 & 1 & -1 \\
0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 \\
\end{pmatrix}.
\]
The rank and nullity theorem has a corollary that says, if $T : \mathbb{R}^n \to \mathbb{R}^m$ is linear, then $T$ is one-to-one if and only if $T$ is onto. Thus if we want to check such a transformation for invertibility, we don’t need to see if it is both one-to-one and onto, we can just check for one of these two properties. A glance at the columns of a $2 \times 2$ matrix is all that is needed to see if the corresponding transformation is onto, and hence invertible. The transformation fails to be invertible only when one of the columns is a multiple of the other. A glance at a $3 \times 3$ matrix might not be enough to determine the invertibility of the corresponding transformation because now we need to see if any column vector is in the plane determined by the other two column vectors. If the glance is not enough, a quick computation surely will be.

**Example 7**

Determine whether the transformations corresponding to the following matrices are invertible.

\[
\begin{align*}
i & \quad \begin{pmatrix} 0 & 1 \\ 1 & 1 \end{pmatrix}_B \\
ii & \quad \begin{pmatrix} 1 & 2 & 3 \\ 8 & 6 & 4 \\ 7 & 6 & 5 \end{pmatrix}_B \\
iii & \quad \begin{pmatrix} 1 & 2 & 3 \\ 8 & 16 & 4 \\ 7 & 14 & 5 \end{pmatrix}_B
\end{align*}
\]

The $2 \times 2$ matrix is invertible because the second column is not a multiple of the first. The first $3 \times 3$ matrix is not so clear; we need to see whether the range is two or three dimensional. The first two columns already span two dimensions inside the range, so if the range were two dimensional the third column vector would have to be in the plane determined from the first two columns; i.e. the equation

\[
\alpha \begin{pmatrix} 1 \\ 8 \\ 7 \end{pmatrix} + \beta \begin{pmatrix} 2 \\ 6 \\ 6 \end{pmatrix} = \begin{pmatrix} 3 \\ 4 \\ 5 \end{pmatrix}
\]

would be solvable. The first two coordinates give the system

\[
\begin{align*}
\alpha + 2\beta &= 3 \\
8\alpha + 6\beta &= 4,
\end{align*}
\]

which completely determines $\alpha$ and $\beta$: $\alpha = -1$ and $\beta = 2$. Do these scalars work in the third coordinate? We check: $7\alpha + 6\beta = (7)(-1) + (6)(2) = 5$, so the third column is in the span of the first two, and the corresponding transformation is not invertible. The rightmost $3 \times 3$ has rank two, and is consequently not invertible, since the second column is a multiple of the first, so the span of the three columns is the same as the span of the first and last column.

**Problem 19** Determine whether the transformations corresponding to the following matrices are invertible:

\[
\begin{align*}
i & \quad \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}_B \\
ii & \quad \begin{pmatrix} 1 & 2 & 3 \\ 8 & 9 & 4 \\ 7 & 6 & 5 \end{pmatrix}_B \\
iii & \quad \begin{pmatrix} 1 & 0 & 3 \\ 8 & 0 & 4 \\ 7 & 0 & 5 \end{pmatrix}_B
\end{align*}
\]

A casual glance at a transformation’s matrix is seldom enough to see what the nullspace is, unless a column of zeros appears, which reveals that one of the basis vectors is in the nullspace. The rightmost matrix in problem 19 proclaims quite loudly that the second basis vector in $\mathcal{B}$ is in the nullspace. Since that matrix has rank two, we conclude that the nullspace is in fact the span of that second basis vector.
3.6 Invariants

We have seen repeatedly how a single transformation can give rise to a multitude of formulae and $B$-matrices, depending on the various choices for the basis $B$. Even though the change in the formulae seems random, it is not, and there are things that do not change; we refer to these as similarity invariants of the transformation, or simply as invariants. One of these invariants is the sum of the entries on the main diagonal; this is called the trace of the matrix. The transformation $T(x, y) = (x, \frac{1}{2}y)$, which figured prominently in the section on vector fields, gives rise to all of the following matrices (depending on the choice of $B$):

$$
\begin{pmatrix}
1 & 0 \\
0 & \frac{1}{2}
\end{pmatrix}_B,
\begin{pmatrix}
\frac{3}{4} & -\frac{1}{4} \\
-\frac{1}{4} & \frac{3}{4}
\end{pmatrix}_B,
\begin{pmatrix}
1 & \pi \\
0 & \frac{1}{2}
\end{pmatrix}_B.
$$

Notice that the trace in each case is 1.5, and the eigenvalues are clearly visible in the diagonal and upper triangular form, while they are completely obscured in the middle matrix.

**Example 8**

Assume $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is given by the formula $T(x, y) = (2x + y, x + 2y)$. Find the standard matrix of $T$ and find the $B$-matrix of $T$ when $B = \{(1,1), (1,-1)\}$.

The standard matrix is obtained by writing $T(1,0)$ in the first column and $T(0,1)$ in the second column, using standard coordinates. Inputting these vectors into the given formula for $T$ yields

$$
\begin{pmatrix}
2 & 1 \\
1 & 2
\end{pmatrix},
$$

which has a trace of 4, and gives no clue about the eigenvalues.

To find the $B$-matrix for $T$, we need to compute $T(1,1)$ and $T(1,-1)$, and put these vectors in the first and second columns (respectively) in terms of their $B$-coordinates. In terms of standard coordinates, we compute $T(1,1) = (3,3)$, which needs to be converted to $B$-coordinates; so write

$$
(3,3) = \alpha(1,1) + \beta(1,-1)
$$

and solve, obtaining $\alpha = 3$ and $\beta = 0$. So far, the solution looks like

$$
\begin{pmatrix}
3 & ? \\
0 & ?
\end{pmatrix}_B,
$$

and without doing anymore work we can already tell that the matrix is upper triangular, and thus 3 is an eigenvalue. But, if we were awake when we performed the computation above, we would have noticed that $T(1,1) = 3(1,1)$, so of course 3 is an eigenvalue.

To get the second column, we compute $T(1,-1) = (1,-1)$, and convert;

$$
(1,-1) = \alpha(1,1) + \beta(1,-1),
$$

to get the $B$-coordinates $(1,-1) = (\alpha, \beta)_B = (0,1)_B$, which we put in the second column to obtain the final answer;

$$
\begin{pmatrix}
3 & 0 \\
0 & 1
\end{pmatrix}_B.
$$

The trace is still 4, which is a good omen.

**Problem 20**

Assume $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is given by the formula $T(x, y) = (x + 2y, 2x + y)$. Find the standard matrix of $T$ and find the $B$-matrix of $T$ when $B = \{(1,1), (1,-1)\}$.
Example 9

Assume \( T: \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) is given by the formula \( T(x, y, z) = (2x + y + z, x - y + z, x + y - z) \). Find the standard matrix of \( T \) and find the \( \mathcal{B} \)-matrix of \( T \) when

\[
\mathcal{B} = \{(1, 1, 1), (1, -1, 1), (1, -1, -1)\}.
\]

The standard matrix is obtained by computing \( T(1, 0, 0), T(0, 1, 0), \) and \( T(0, 0, 1) \), then writing these vectors in the columns using their standard coordinates: for example, we compute

\[
T(1, 0, 0) = (2, 1, 1)
\]

to see that the entries of the first column are 2, 1, and 1. Continuing in this way, we arrive at

\[
\begin{pmatrix}
2 & 1 & 1 \\
1 & -1 & 1 \\
1 & 1 & -1
\end{pmatrix}
\]

as the standard matrix for \( T \), which has trace 0.

To get the \( \mathcal{B} \)-matrix of \( T \), we compute \( T(1, 1, 1), T(1, -1, 1), \) and \( T(1, -1, -1) \), then convert these vectors into their \( \mathcal{B} \)-coordinates before writing them in the columns. The formula for \( T \) tells us that \( T(1, 1, 1) = (4, 1, 1) \), and the \( \mathcal{B} \)-coordinates of this vector are obtained by writing

\[
(4, 1, 1) = \alpha (1, 1, 1) + \beta (1, -1, 1) + \gamma (1, -1, -1),
\]

and solving for \( \alpha \), \( \beta \), and \( \gamma \). This equation is equivalent to the following system of three equations in three unknowns,

\[
\begin{align*}
\alpha + \beta + \gamma &= 4 \\
\alpha - \beta - \gamma &= 1 \\
\alpha + \beta - \gamma &= 1,
\end{align*}
\]

which has the solution \( \alpha = \frac{5}{2}, \beta = 0, \) and \( \gamma = \frac{3}{2} \), and these three numbers go in the first column of the \( \mathcal{B} \)-matrix. To get the second column, we are required to compute \( T(1, -1, 1) = (2, 3, -1) \), and convert to \( \mathcal{B} \)-coordinates

\[
(2, 3, -1) = \alpha (1, 1, 1) + \beta (1, -1, 1) + \gamma (1, -1, -1),
\]

leading to the system

\[
\begin{align*}
\alpha + \beta + \gamma &= 2 \\
\alpha - \beta - \gamma &= 3 \\
\alpha + \beta - \gamma &= -1,
\end{align*}
\]

which has the solution \( \alpha = \frac{5}{2}, \beta = -2, \) and \( \gamma = \frac{3}{2} \). So far, the the \( \mathcal{B} \)-matrix for \( T \) looks like

\[
\begin{pmatrix}
\frac{5}{2} & \frac{5}{2} & ? \\
0 & -2 & ? \\
\frac{5}{2} & \frac{3}{2} & ?
\end{pmatrix}_{\mathcal{B}}.
\]

If we haven’t made a mistake yet, we are going to get \( -\frac{1}{2} \) in the bottom right; otherwise the trace wouldn’t add to 0. Let’s compute: \( \ldots T(1, -1, -1) = (0, 1, 1) \ldots \)

\[
(0, 1, 1) = \alpha (1, 1, 1) + \beta (1, -1, 1) + \gamma (1, -1, -1),
\]

leading to the system

\[
\begin{align*}
\alpha + \beta + \gamma &= 0 \\
\alpha - \beta - \gamma &= 1 \\
\alpha + \beta - \gamma &= -1,
\end{align*}
\]

which has the solution \( \alpha = \frac{5}{2}, \beta = -2, \) and \( \gamma = \frac{3}{2} \). Thus the \( \mathcal{B} \)-matrix for \( T \) is

\[
\begin{pmatrix}
\frac{5}{2} & \frac{5}{2} & \frac{1}{2} \\
0 & -2 & \frac{1}{2} \\
\frac{5}{2} & \frac{3}{2} & \frac{1}{2}
\end{pmatrix}_{\mathcal{B}}.
\]
\[
\begin{align*}
\alpha + \beta + \gamma &= 0 \\
\alpha - \beta - \gamma &= 1 \\
\alpha + \beta - \gamma &= 1,
\end{align*}
\]
\[
\begin{align*}
\cdots \alpha &= \frac{1}{2}, \beta = 0, \text{ and } \gamma = -\frac{1}{2} \cdots \\
\begin{pmatrix} 5 \\ 0 \\ 3 \\ 2 \\ \frac{1}{2} \\ 0 \end{pmatrix}
\end{align*}
\]
which still has trace 0, so there’s a good chance this answer is right.

**Problem 21** Assume \( T : \mathbb{R}^3 \to \mathbb{R}^3 \) is given by the formula \( T(x, y, z) = (x + y + 2z, x + z, 2x + y - z) \). Find the standard matrix of \( T \) and find the \( B \)-matrix of \( T \) when \( B = \{(1, 1, 1), (1, -1, 1), (1, -1, -1)\} \).

Use the traces of the matrices to check your answers.

Another similarity invariant is the **determinant** of the matrix: for a \( 2 \times 2 \) matrix
\[
\begin{pmatrix} a & b \\ c & d \end{pmatrix},
\]
the determinant is \( ad - bc \). For the \( 3 \times 3 \) matrix
\[
\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix},
\]
the determinant is
\[
 ace - afh - bdi + bfg + cdh - cge.
\]
The determinant is a measure of something intrinsic about the transformation; for a transformation \( T : \mathbb{R}^2 \to \mathbb{R}^2 \), the determinant of any \( B \)-matrix for \( T \) gives the signed area of the image of the unit square (see figure ??). The unit square is the parallelogram determined by \((1,0)\) and \((0,1)\), and the sign of the determinant reflects whether the edges of the parallelogram determined by \(T(1,0)\) and \(T(0,1)\) are in the same relationship as the edges of the unit square. During the transformation of the unit square, the parallelogram might flip, revealing its bottom, and the determinant would be negative. If no flip occurs, the determinant is positive. The absolute value of the determinant is the area of the parallelogram determined by \(T(1,0)\) and \(T(0,1)\); since we started with the unit square, which has area 1, the determinant gives an exact measure of how much the transformation compresses or expands \( \mathbb{R}^2 \). For functions \( T : \mathbb{R}^n \to \mathbb{R}^n \), the determinant reveals the same type of information about \( T \), but it quickly becomes difficult to compute; if \( T \) has a corresponding \( B \)-matrix \((a_{ij})_B\), the determinant’s general definition is
\[
\sum_{\sigma} \text{sgn}(\sigma)a_{1\sigma_1}a_{1\sigma_2}\cdots a_{n\sigma_n},
\]
where the sum is taken over all permutations \( \sigma \) of the set \( \{1, 2, \ldots, n\} \). Those who know about permutations will realize that such a sum has \( n! \) summands, which quickly becomes cumbersome, even for the fastest computers.
Example 10

Let $C = \{(1, 1), (1, 0)\}$ and assume $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the transformation whose $C$-matrix is

$$
\begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}_C.
$$

With $B = \{(1, 0), (-1, -1)\}$, find the $B$-matrix of $T$ and check the answer by computing determinants.

We need to compute $T(1, 0)$ and $T(-1, -1)$, then put the results in a matrix in terms of these vector’s $B$-coordinates. The problem is, we only know the $C$-matrix for $T$ and the vectors $(1, 0)$ and $(-1, -1)$ are given in standard coordinates. We could proceed in one of two ways; we could find the standard matrix of $T$ and use it to compute $T(1, 0)$ and $T(-1, -1)$, or we could convert the inputs into $C$-coordinates and use the $C$-matrix for the computations. We will do the latter; so, begining with $(1, 0)$, we seek its $C$-coordinates by writing

$$(1, 0) = \alpha(1, 1) + \beta(1, 0),$$

then solving for $\alpha$ and $\beta$. We get $\alpha = 0$ and $\beta = 1$, and $(1, 0) = (0, 1)_C$. We obtain the $C$-coordinates of (-1,-1) by writing

$$(-1, -1) = \alpha(1, 1) + \beta(1, 0),$$

and solving for $\alpha$ and $\beta$, obtaining $(-1, -1) = (-1, 0)_C$

Now we use the $C$-matrix to see where these two vectors are transformed:

$$
\begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}_C
\begin{pmatrix}
0 \\
1
\end{pmatrix}_C =
\begin{pmatrix}
1 \\
1
\end{pmatrix}_C 	ext{ and } 
\begin{pmatrix}
1 & 1 \\
1 & 1
\end{pmatrix}_C
\begin{pmatrix}
-1 \\
0
\end{pmatrix}_C =
\begin{pmatrix}
-1 \\
-1
\end{pmatrix}_C
$$

It’s convenient that these outputs are negatives of each other, because it saves us a step; all we need to do now is find the $B$-coordinates of the vector $(1, 1)_C$, put those coordinates in the first column, and put the negatives in the second column. We write

$$(1, 1)_C = \alpha(0, 1)_C + \beta(-1, 0)_C,$$

and solve for $\alpha$ and $\beta$, getting $\alpha = 1$ and $\beta = -1$. Thus the $B$-matrix of $T$ is

$$
\begin{pmatrix}
1 & -1 \\
1 & -1
\end{pmatrix}_B.
$$

The determinants of both the $B$-matrix and the $C$-matrix are 0, so perhaps we didn’t make any mistakes.

When we worked to obtain the $B$-coordinates of $(1, 1)_C$, we did so using $C$-coordinates. Notice that, in terms of $C$-coordinates, we have $B = \{(0, 1)_C, (-1, 0)_C\}$. We could have done the same computation in standard coordinates: the standard coordinates of $(1, 1)_C$ are $(2, 1)$, and computing the $B$-coordinates involves solving the system

$$(2, 1) = \alpha(1, 0) + \beta(-1, -1),$$

which again gives us $\alpha = 1$ and $\beta = -1$. The computation can be done in any coordinate system. The only hazzard is mixing coordinate systems in a computation; don’t do this, because if you do, only good luck will turn out the correct answer.
Problem 22  Let $\mathcal{C} = \{(1, 1), (1, 0)\}$ and assume $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the transformation whose $\mathcal{C}$-matrix is

$$\begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}_c.$$  

With $\mathcal{B} = \{(1, 0), (-1, -1)\}$, find the $\mathcal{B}$-matrix of $T$ and check the answer by computing determinants.

If a transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ has a $\mathcal{B}$-matrix

$$\begin{pmatrix} a & b \\ c & d \end{pmatrix}_\mathcal{B},$$

then the determinant of

$$\begin{pmatrix} x & 0 \\ 0 & x \end{pmatrix}_\mathcal{B} - \begin{pmatrix} a & b \\ c & d \end{pmatrix}_\mathcal{B} = \begin{pmatrix} x-a & -b \\ -c & x-d \end{pmatrix}_\mathcal{B}$$

is called the characteristic polynomial of $T$, and it is another similarity invariant. For a transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ with a $\mathcal{B}$-matrix

$$\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}_\mathcal{B},$$

the characteristic polynomial is the determinant of

$$\begin{pmatrix} x & 0 & 0 \\ 0 & x & 0 \\ 0 & 0 & x \end{pmatrix}_\mathcal{B} - \begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix}_\mathcal{B} = \begin{pmatrix} x-a & -b & -c \\ -d & x-e & -f \\ -g & -h & x-i \end{pmatrix}_\mathcal{B}.$$  

For general transformations $T : \mathbb{R}^n \to \mathbb{R}^n$, the characteristic polynomial is the determinant of

$$xI - T,$$

where $I$ denotes the identity transformation on $\mathbb{R}^n$. This characteristic polynomial has degree $n$, and the roots of the polynomial are exactly the eigenvalues of $T$.

The calculation of determinants (and, in particular, of characteristic polynomials) is greatly simplified if one first goes to the trouble of finding an upper (or lower) triangular form for the transformation. When the matrix has one of these triangular forms, the determinant reduces to the product of the entries on the main diagonal. Thus the determinant of

$$\begin{pmatrix} 3 & 7 \\ 0 & 4 \end{pmatrix}_\mathcal{B}$$

is 12, and the characteristic polynomial is $(x - 3)(x - 4)$. The determinant of

$$\begin{pmatrix} 0 & 0 & 0 \\ \pi & 5 & 0 \\ 2 & 8 & 7 \end{pmatrix}_\mathcal{B}$$

is 0, and the characteristic polynomial is $(x - 0)(x - 5)(x - 7)$.

Example 11  Assume $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the transformation that rotates vectors $90^\circ$ counterclockwise. Find the standard matrix of $T$, and find the $\mathcal{B}$-matrix of $T$ when $\mathcal{B} = \{(1, 1), (1, 2)\}$. Check your answer by computing characteristic polynomials.
To find the standard matrix of $T$, we see where $T$ sends the standard basis vectors $(1, 0)$ and $(0, 1)$. If we rotate $(1, 0)$ by $90^\circ$ counterclockwise, we get $(0, 1)$, so this is the first column. If we rotate $(0, 1)$ by $90^\circ$ counterclockwise, we get $(-1, 0)$, and the standard matrix is

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

The characteristic polynomial of this matrix is the determinant of

$$ \begin{pmatrix} x & 1 \\ -1 & x \end{pmatrix}, $$

which is $x^2 + 1$.

To get the $\mathcal{B}$-matrix of $T$ we need to compute $T(1, 1)$ and $T(1, 2)$, then write these vectors in the columns of the matrix using their $\mathcal{B}$-coordinates. We can use the standard matrix to make the computations:

$$ \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ 1 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \end{pmatrix} = \begin{pmatrix} -2 \\ 1 \end{pmatrix}. $$

Thus $T(1, 1) = (-1, 1)$ and $T(1, 2) = (2, 1)$; converting the first of these to $\mathcal{B}$-coordinates, we write

$$ (-1, 1) = \alpha(1, 1) + \beta(1, 2) $$

and solve, getting $\alpha = -3$ and $\beta = 2$, so $T(1, 1) = (-3, 2)_B$. Converting the second vector, we write

$$ (-2, 1) = \alpha(1, 1) + \beta(1, 2) $$

and solve, getting $\alpha = -5$ and $\beta = 3$, so $T(1, 1) = (-5, 3)_B$, and the $\mathcal{B}$-matrix of $T$ is

$$ \begin{pmatrix} -3 & -5 \\ 2 & 3 \end{pmatrix}_B. $$

The characteristic polynomial is the determinant of

$$ \begin{pmatrix} x + 3 & 5 \\ -2 & x - 3 \end{pmatrix}_B, $$

which is $(x + 3)(x - 3) - (-2)(5) = x^2 + 1$, and now we are confident that we made no mistake.

**Problem 23** Assume $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is the transformation that reflects vectors across the $y$-axis. Find the standard matrix of $T$, and find the $\mathcal{B}$-matrix of $T$ when $\mathcal{B} = \{(1, 1), (1, 2)\}$. Check your answer by computing characteristic polynomials.

If $T$ denotes either a transformation or one of its $\mathcal{B}$-matrices, we let $\text{tr}(T)$ denote its trace and $\det(T)$ denote its determinant. Of the three invariants used for checking the accuracy of our change of coordinates, the characteristic polynomial is the best because it incorporates the other two. In fact, for a transformation $T$ with matrix

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

the characteristic polynomial is

$$(x - a)(x - d) - cb = x^2 - \text{tr}(T)x + \det(T).$$
Example 12

Assume \( T : \mathbb{C}^2 \to \mathbb{C}^2 \) is the transformation whose standard matrix is

\[
\begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix}.
\]

Find the \( B \)-matrix of \( T \) when \( B = \{(i, 1), (1, i)\} \). Check your answer by computing characteristic polynomials.

We will use the standard matrix to compute \( T(i, 1) \) and \( T(1, i) \);

\[
\begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix} \begin{pmatrix}
i \\
1
\end{pmatrix} = \begin{pmatrix} 1 \\
-i
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
0 & 1 \\
-1 & 0
\end{pmatrix} \begin{pmatrix}
i \\
1
\end{pmatrix} = \begin{pmatrix} i \\
-1
\end{pmatrix},
\]

so \( T(i, 1) = (1, -i) \) and \( T(1, i) = (i, -1) \). Converting these to \( B \)-coordinates, beginning with \( T(i, 1) \), we write

\[
(1, -i) = \alpha(i, 1) + \beta(1, i)
\]

and solve to get \( \alpha = -i \) and \( \beta = 0 \). So far, our \( B \)-matrix looks like

\[
\begin{pmatrix}
-i & ? \\
0 & ?
\end{pmatrix}_B.
\]

Converting \( T(1, i) \) into \( B \)-coordinates, we write

\[
(i, -1) = \alpha(i, 1) + \beta(1, i)
\]

and solve to get \( \alpha = 0 \) and \( \beta = i \). Thus the \( B \)-matrix is

\[
\begin{pmatrix}
-i & 0 \\
0 & i
\end{pmatrix}_B,
\]

which has the characteristic polynomial \((x + i)(x - i)\), which is the same as \( x^2 + 1 \).

Problem 24

Assume \( T : \mathbb{C}^2 \to \mathbb{C}^2 \) is the transformation whose standard matrix is

\[
\begin{pmatrix}
0 & i \\
i & 0
\end{pmatrix}.
\]

Find the \( B \)-matrix of \( T \) when \( B = \{(1, i), (-1, i)\} \). Check your answer by computing characteristic polynomials.

More Exercises

1. Use the rank and nullity theorem to prove the following: if \( T : \mathbb{R}^n \to \mathbb{R}^n \) is linear then \( T \) is one-to-one if and only if \( T \) is onto.
2. Show that the system

\[
\begin{align*}
ax + by &= e \\
x + dy &= f
\end{align*}
\]

has a solution when \( ad - bc \neq 0 \).
3. Show that if the system

\[
\begin{align*}
ax + by &= e \\
x + dy &= f
\end{align*}
\]

has a solution, then \( ad - bc \neq 0 \).
Chapter 4
Diagonalization

4.1 An algorithm

The idea is not to randomly change coordinates and watch the matrices change with them, as we have asked the reader to do several times in the problems. We want the reader to have this experience so they would more fully appreciate the true crucial idea, which is to carefully choose a coordinate system that gives rise to a diagonal matrix form. When this can be done, all of the eigenvalues are exhibited in the matrix, and the corresponding eigenvectors constitute the basis. What we need is an algebraic method of finding invariant subspaces for a transformation, and that is what this section delivers. We already have introduced the most important player of this algebraic method, which is the characteristic polynomial.

The thing to remember about the characteristic polynomial is that its roots are the eigenvalues of the transformation. Thus if we are given a transformation $T : R^2 \to R^2$, we can find its eigenvalues by using the quadratic formula on its characteristic polynomial.

Example 1

Find the eigenvalues for the transformation $T : R^2 \to R^2$ with $\mathcal{B}$-matrix

$$
\begin{pmatrix}
0 & -6 \\
1 & 5
\end{pmatrix}_B.
$$

The characteristic polynomial is defined to be the determinant of

$$
x\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}_B - \begin{pmatrix}
0 & -6 \\
1 & 5
\end{pmatrix}_B = \begin{pmatrix}
x & 6 \\
-1 & x - 5
\end{pmatrix}_B,
$$

which is $x(x - 5) - (-1)(6) = x^2 - 5x + 6$, which has roots 3 and 2 (which can be found via the quadratic formula, if they can not be seen with the naked eye). Thus the eigenvalues of $T$ are 3 and 2.

Problem 1

Find the eigenvalues for the transformation $T : R^2 \to R^2$ with $\mathcal{B}$-matrix

$$
\begin{pmatrix}
0 & 3 \\
1 & -2
\end{pmatrix}_B.
$$

If we know that $T$ has 3 for an eigenvalue, we can find the invariant line corresponding to 3 by solving the equation $T(u) = 3u$ for the unknown $u$. If we bring everything to the right, this equation looks like

$$
o = (3I - T)u,
$$
which makes solving the equation the same as finding a nullspace.

**Example 2**

Find the eigenvectors for the transformation \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) with \( B \)-matrix

\[
\begin{pmatrix}
0 & -6 \\
1 & 5 \\
\end{pmatrix}_B.
\]

This is the same transformation that we met in example 1, where we saw that 2 and 3 are eigenvalues. Beginning with the eigenvalue 3, we seek the nullspace of \( 3I - T \), i.e. the nullspace corresponding to the matrix used to calculate the characteristic polynomial of \( T \), with 3 plugged in for \( x \). That matrix is

\[
\begin{pmatrix}
x & 6 \\
-1 & x - 5 \\
\end{pmatrix}_B
\]

and replacing \( x \) with 3 results in

\[
\begin{pmatrix}
3 & 6 \\
-1 & -2 \\
\end{pmatrix}_B.
\]

The nullspace can be determined by looking at the dependency in the columns; the second column is twice the first column, so \( (2, -1)_B \) is in the nullspace by the way we defined the multiplication:

\[
\begin{pmatrix}
3 & 6 \\
-1 & -2 \\
\end{pmatrix}_B \begin{pmatrix}
2 \\
-1 \\
\end{pmatrix}_B = 2 \begin{pmatrix}
3 \\
-1 \\
\end{pmatrix}_B + (-1) \begin{pmatrix}
6 \\
-2 \\
\end{pmatrix}_B = \begin{pmatrix}
0 \\
0 \\
\end{pmatrix}_B.
\]

Thus the span of \( (2, -1)_B \) is the invariant line corresponding to the eigenvalue 3; everything on this line is transformed into 3 times itself, and the nonzero vectors on this line are the eigenvectors corresponding to 3.

To get the invariant line corresponding to the eigenvalue 2, we plug 2 in for \( x \) in the matrix

\[
\begin{pmatrix}
x & 6 \\
-1 & x - 5 \\
\end{pmatrix}_B.
\]

When we do, we get the matrix

\[
\begin{pmatrix}
2 & 6 \\
-1 & -3 \\
\end{pmatrix}_B
\]

for which the second column is 3 times the first, so \( (3, -1)_B \) is in its nullspace since

\[
\begin{pmatrix}
2 & 6 \\
-1 & -3 \\
\end{pmatrix}_B \begin{pmatrix}
3 \\
-1 \\
\end{pmatrix}_B = 3 \begin{pmatrix}
2 \\
-1 \\
\end{pmatrix}_B + (-1) \begin{pmatrix}
6 \\
-3 \\
\end{pmatrix}_B = \begin{pmatrix}
0 \\
0 \\
\end{pmatrix}_B.
\]

Thus the line spanned by \( (3, -1)_B \) is an invariant line, on which the nonzero vectors are eigenvectors corresponding to the eigenvalue 2. Every vector on this line is transformed into twice itself.

**Problem 2**

Find the eigenvectors for the transformation \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) with \( B \)-matrix

\[
\begin{pmatrix}
0 & 3 \\
1 & -2 \\
\end{pmatrix}_B.
\]

**Example 3**

Find a basis \( B \) consisting of eigenvectors for the transformation \( T : \mathbb{R}^2 \to \mathbb{R}^2 \) whose standard matrix is

\[
\begin{pmatrix}
2 & -1 \\
5 & 8 \\
\end{pmatrix},
\]

and give the corresponding \( B \)-matrix of \( T \).
We first look for the eigenvalues of $T$; the determinant of
\[
\begin{pmatrix}
  x - 2 & 1 \\
  -5 & x - 8
\end{pmatrix}
\]
is $(x - 2)(x - 8) - (5)(-1) = x^2 - 10x + 21$, which has roots $x = 3$ and $x = 7$. Plugging 3 in for $x$ into the matrix above, we get
\[
\begin{pmatrix}
  1 & 1 \\
  -5 & -5
\end{pmatrix},
\]
whose nullspace is spanned by $(1, -1)$. Any nonzero vector on this line is an eigenvector corresponding to the eigenvalue 3, and we are free to select any one of them for our basis; let’s take $(1, -1)$. So far, our basis looks like $\mathbf{B} = \{(1, -1), (? , ?)\}$ and the $\mathbf{B}$-matrix looks like
\[
\begin{pmatrix}
  3 & ? \\
  0 & ?
\end{pmatrix}_{\mathbf{g}}.
\]
To find a second eigenvector independent of the first, plug $x = 7$ into the matrix above to get
\[
\begin{pmatrix}
  5 & 1 \\
  -5 & -1
\end{pmatrix},
\]
and find a nonzero vector in the nullspace, like $(1, -5)$. This vector is then an eigenvector for $T$ corresponding to the eigenvalue 7, and $\mathbf{B} = \{(1, -1), (1, -5)\}$ is a basis consisting of eigenvectors for $T$. The corresponding $\mathbf{B}$-matrix is
\[
\begin{pmatrix}
  3 & 0 \\
  0 & 7
\end{pmatrix}_{\mathbf{g}}.
\]

**Problem 3** Find a basis $\mathbf{B}$ consisting of eigenvectors for the transformation $T : \mathbb{R}^2 \to \mathbb{R}^2$ whose standard matrix is
\[
\begin{pmatrix}
  2 & -1 \\
  8 & 8
\end{pmatrix},
\]
and give the corresponding $\mathbf{B}$-matrix of $T$.

**Example 4** Assume that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the transformation that sends the vector $\mathbf{u}$ in figure 1 to the vector $\mathbf{v}$ and sends $\mathbf{v}$ to $\mathbf{u}$. Find a basis $\mathbf{B}$ consisting of eigenvectors for $T$, and give the corresponding $\mathbf{B}$-matrix.

![Fig. 1](image)

We begin by looking for any matrix for $T$; since we know what $T$ does to $\mathbf{u}$ and $\mathbf{v}$, let $\mathcal{C} = \{\mathbf{u}, \mathbf{v}\}$ and write the $\mathcal{C}$-matrix of $T$. The first column of this matrix is the vector $T(\mathbf{u})$ written in $\mathcal{C}$-coordinates, i.e it is $\mathbf{v}$ in $\mathcal{C}$ coordinates; thus the first column of the $\mathcal{C}$-matrix of $T$ looks like
\[
\begin{pmatrix}
  0 & ? \\
  1 & ?
\end{pmatrix}_{\mathcal{C}}.
\]
The second column contains the $\mathcal{C}$-coordinates of $T(\mathbf{v}) = \mathbf{u}$, which are $(1, 0)_{\mathcal{C}}$, and the $\mathcal{C}$-matrix of $T$ is
\[
\begin{pmatrix}
  0 & 1 \\
  1 & 0
\end{pmatrix}_{\mathcal{C}}.
\]
Now we perform the diagonalization algorithm on this matrix. The characteristic polynomial is the determinant of
\[
\begin{pmatrix}
  x & -1 \\
  -1 & x
\end{pmatrix}_{\mathcal{C}},
\]
which is \(x^2 - 1\), so the eigenvalues are 1 and –1. Plugging \(x = 1\) into the matrix above, we have

\[
\begin{pmatrix}
1 & -1 \\
-1 & 1
\end{pmatrix}_c,
\]

whose nullspace is spanned by \((1, 1)_c = u + v\). Thus \(u + v\) is an eigenvector for \(T\). Plugging \(x = -1\) in above yields a second independent eigenvector: the nullspace of

\[
\begin{pmatrix}
-1 & -1 \\
-1 & -1
\end{pmatrix}_c
\]

is spanned by \((1, -1)_c = u - v\). We build a basis using these two eigenvectors, declaring that \(B = \{u + v, u - v\}\), and the corresponding \(B\)-matrix for \(T\) is

\[
\begin{pmatrix}
1 & 0 \\
0 & -1
\end{pmatrix}_B.
\]

**Problem 4** Assume that \(T : R^2 \rightarrow R^2\) is the transformation that sends the vector \(u\) in figure 2 to the vector \(w\) and sends \(v\) to \(w\). Find a basis \(B\) consisting of eigenvectors for \(T\), and give the corresponding \(B\)-matrix.

### 4.2 The fundamental theorem of algebra

A transformation \(T : R^2 \rightarrow R^2\) might fail to have any eigenvectors or eigenvalues. The 90° counterclockwise rotation has no invariant lines, hence no eigenvectors. This geometric observation has an algebraic counterpart: the characteristic polynomial of such a rotation has no real roots, hence no eigenvalues. The principle advantage of using complex scalars arises as a consequence of the **fundamental theorem of algebra**, which says that every nonconstant complex polynomial has a root. The familiar quadratic formula

\[
-b \pm \sqrt{b^2 - 4ac}
\]

\[
2a
\]

gives the roots of a quadratic polynomial \(ax^2 + bx + c\), and when the discriminant (the quantity within the square root) is negative, the roots are complex. The ability to find an eigenvalue, and hence a corresponding eigenvector, guarantees that a transformation of \(C^2\), or more generally, any transformation of \(C^n\), has a triangular matrix form, and many transformations that are not diagonalizable using real scalars become so when using complex scalars.

**Example 5**

Assume \(T : C^2 \rightarrow C^2\) has the standard matrix

\[
\begin{pmatrix}
0 & -2 \\
1 & 2
\end{pmatrix}.
\]

Find a basis \(B\) consisting of eigenvectors of \(T\), and give the corresponding \(B\)-matrix.

The characteristic polynomial of \(T\) is the determinant of

\[
\begin{pmatrix}
x & 2 \\
-1 & x - 2
\end{pmatrix},
\]
which is \( x^2 - 2x + 2 \). The quadratic formula tells us that the roots are

\[
\frac{2 \pm \sqrt{2^2 - 8}}{2},
\]

i.e. the roots are \( 1 + i \) and \( 1 - i \). Let \( x = 1 + i \) in the matrix above to get

\[
\begin{pmatrix}
1 + i & 2 \\
-1 & i - 1
\end{pmatrix},
\]

in order to find a nonzero vector in the nullspace of this matrix, we need to see a dependency between the columns. If the second column is a multiple of the first, the \(-1\) and \( i - 1 \) entries in the second row will tell us what that multiple is; \( \alpha (-1) = i - 1 \) happens only when \( \alpha = 1 - i \), so we check to see if

\[
(1 - i) \begin{pmatrix} 1 + i \\ -1 \end{pmatrix} \neq \begin{pmatrix} 2 \\ i - 1 \end{pmatrix},
\]

and we are satisfied, since \((1 - i)(1 + i) = 2\). Knowing this dependency tells us that \((1 - i, -1)\) is in the nullspace of the matrix, so this is an eigenvector of \( T \) corresponding to the eigenvalue \( 1 + i \).

In search of an eigenvector to correspond with \( 1 - i \), we plug \( x = 1 - i \) into the matrix above to get

\[
\begin{pmatrix}
1 - i & 2 \\
-1 & -i - 1
\end{pmatrix},
\]

and the second row suggests that \((i + 1)\) times the first column equals the second. We verify that this multiple works for the entries in the first row, which again boils down to \((1 - i)(1 + i) = 2\). Thus \((i + 1, -1)\) is an eigenvector corresponding to the eigenvalue \( 1 - i \). Putting these two eigenvectors into a basis \( B = \{(1 - i, -1), (i + 1, -1)\} \), we obtain a basis consisting of eigenvectors for which the corresponding \( B \)-matrix of \( T \) is

\[
\begin{pmatrix}
1 + i & 0 \\
0 & 1 - i
\end{pmatrix}_B.
\]

### Problem 5

Assume \( T : C^2 \to C^2 \) has the standard matrix

\[
\begin{pmatrix}
0 & -5 \\
1 & 4
\end{pmatrix}.
\]

Find a basis \( B \) consisting of eigenvectors of \( T \), and give the corresponding \( B \)-matrix.

For transformations \( T \) of a two dimensional space, if the characteristic polynomial has two distinct roots, then an eigenvector can be obtained for each of these eigenvalues, and that pair of eigenvectors forms a basis that diagonalizes \( T \). On the other hand, if the characteristic polynomial has only one root, like the polynomial \( x^2 - 2x + 1 \), then the root is repeated and we refer to the repetition as the root’s **multiplicity**. There is an example of a diagonalizable transformation \( T \) whose sole eigenvalue has multiplicity; being diagonalizable with a single eigenvalue \( \alpha \) means that the transformation has a diagonal matrix form

\[
\begin{pmatrix}
\alpha & 0 \\
0 & \alpha
\end{pmatrix}_B,
\]

which means that this is a **scalar transformation**, a transformation that takes every vector to \( \alpha \) times that vector.

### Problem 6

Assume \( T : C^2 \to C^2 \) is the scalar transformation that sends every vector to \( \alpha \) time that vector; i.e. every nonzero vector is an eigenvector corresponding to the same eigenvalue \( \alpha \). Let \( C = \{u, v\} \) denote an arbitrary basis of \( C^2 \). Show that the \( C \)-matrix of \( T \) is

\[
\begin{pmatrix}
\alpha & 0 \\
0 & \alpha
\end{pmatrix}_C.
\]
When someone says that “the trace of a matrix is the sum of the eigenvalues, including multiplicity, and the determinant is the product of the eigenvalues, including multiplicity” they mean that, when adding or multiplying the eigenvalues, a single eigenvalue should be added or multiplied with itself the same number of times that it occurs as a root. The only obstruction to diagonalization for transformations of complex vector spaces happens when an eigenvalue has multiplicity. Even when we can not diagonalize a transformation of $\mathbb{C}^2$, we can find a triangular matrix form, which is something that can not be said for transformations of $\mathbb{R}^2$. 

**Example 6**

Assume $T : \mathbb{C}^2 \rightarrow \mathbb{C}^2$ is the transformation with standard matrix

\[
\begin{pmatrix}
0 & -1 \\
1 & 2
\end{pmatrix},
\]

Find a basis $\mathcal{B}$ so that the $\mathcal{B}$-matrix of $T$ is upper triangular, and display the resulting $\mathcal{B}$-matrix.

This type of problem is easier than the diagonalizing problems because all we need to do is find one eigenvector. When that eigenvector is used as the first vector in a basis, then the matrix form will be upper triangular no matter what the second basis vector is. To find that eigenvector, we proceed as if we are diagonalizing $T$; the characteristic polynomial of $T$ is the determinant of

\[
\begin{pmatrix}
x & 1 \\
-1 & x - 2
\end{pmatrix},
\]

which is $x^2 - 2x + 1$, and this polynomial has only one root, namely $x = 1$. Plugging this value into the matrix above, we have

\[
\begin{pmatrix}
1 & 1 \\
-1 & -1
\end{pmatrix},
\]

and see that $(1, -1)$ is a vector in the nullspace, and hence an eigenvector for $T$. We build a basis $\mathcal{B}$ by letting $(1, -1)$ be the first basis vector, and randomly throwing in the first vector to come to mind as the second basis vector (the only hazzard is if that random second choice is a multiple of the first vector, which we avoid, since the resulting $\mathcal{B}$ would not then be a basis). So how does $(3, 2)$ look as a second basis vector? We then have a basis $\mathcal{B} = \{(1, -1), (3, 2)\}$, and the $\mathcal{B}$-coordinates of $T(1, -1)$ are $(1, 0)$ because

\[
T(1, -1) = (1, -1) = (1)(1, -1) + (0)(3, 2),
\]

which gives us the first column of the $\mathcal{B}$-matrix. The second column is obtained by computing $T(3, 2)$ and converting the result into $\mathcal{B}$-coordinates; $T(3, 2) = (-2, 7)$ and

\[
(-2, 7) = (-5)(1, -1) + (1)(3, 2),
\]

so the $\mathcal{B}$-coordinates of $T(3, 2)$ are $(-5, 1)$, which give the second column and we have our $\mathcal{B}$-matrix

\[
\begin{pmatrix}
1 & -5 \\
0 & 1
\end{pmatrix}_\mathcal{B}.
\]

Here are a couple of computational observations that are handy to keep in mind. If we let an eigenvector be the first element of a basis, then the first column of the corresponding matrix will always look like the eigenvalue above zeros. When computing the $\mathcal{B}$-coordinates of $T(3, 2)$ above, we did not have to go to the trouble of solving the system

\[
\begin{align*}
\alpha + 3\beta &= -2 \\
-\alpha + 2\beta &= 7
\end{align*}
\]
because we knew the matrix form had to end up looking like
\[
\begin{pmatrix}
1 & ? \\
0 & 1
\end{pmatrix},
\]
because 1 is the only eigenvalue. Thus what we did was let \( \beta = 1 \) and then check to see what \( \alpha \) is in the equation
\[
(-2, 7) = \alpha(1, -1) + (1)(3, 2).
\]

**Problem 7** Assume \( T : \mathbb{C}^2 \rightarrow \mathbb{C}^2 \) is the transformation with standard matrix
\[
\begin{pmatrix}
0 & -4 \\
1 & 4
\end{pmatrix}.
\]
Find a basis \( B \) so that the \( B \)-matrix of \( T \) is upper triangular, and display the resulting \( B \)-matrix.

### 4.3 Eigenspaces

A transformation \( T : \mathbb{C}^3 \rightarrow \mathbb{C}^3 \) might have three distinct eigenvalues, in which case we could find the three corresponding eigenvectors, put them in a basis, and see the resulting diagonalization. If, on the other hand, one of the eigenvalues occurs with multiplicity, then either \( T \) will fail to be diagonalizable, or \( T \) will have a matrix of the form
\[
\begin{pmatrix}
\alpha & 0 & 0 \\
0 & \alpha & 0 \\
0 & 0 & \beta
\end{pmatrix},
\]
In case \( T \) is diagonalizable, then \( T \) behave like a scalar transformation on a two dimensional subspace, namely the space spanned by the first two basis vectors of \( B \). If someone ensures us that a particular transformation \( T : \mathbb{C}^3 \rightarrow \mathbb{C}^3 \) is diagonalizable, and that this \( T \) has only the 2 eigenvalues, then we are going to find that the eigenvectors corresponding to one of the eigenvalues occupy a line, and the eigenvectors corresponding to the other eigenvalue occupy a plane. In the diagonalization algorithm, this is reflected in the step when we plug the eigenvalue in for \( x \) in a matrix form for \( xI - T \); one of the nullspaces will be one dimensional, and the other will be two dimensional. When a transformation has an eigenvalue \( \alpha \), we refer to the nullspace of \( \alpha I - T \) as the **eigenspace** of \( T \) corresponding to \( \alpha \). This eigenspace is a subspace, and this subspace is characterized by writing
\[
\{ v : T(v) = \alpha v \}.
\]

**Example 7**

Assume \( T : \mathbb{C}^3 \rightarrow \mathbb{C}^3 \) is the transformation with standard matrix
\[
\begin{pmatrix}
1 & 1 & 1 \\
0 & 2 & 2 \\
0 & 0 & 3
\end{pmatrix}.
\]
Find a basis \( B \) consisting of eigenvectors, and give the corresponding \( B \)-matrix.

We can see from the triangular form of the standard matrix that the characteristic polynomial is \( (x - 1)(x - 2)(x - 3) \) and the eigenvalues are 1, 2, and 3. We plug each of these in for \( x \) in the standard matrix of \( xI - T \), i.e. in the matrix
\[
\begin{pmatrix}
x - 1 & -1 & -1 \\
0 & x - 2 & -2 \\
0 & 0 & x - 3
\end{pmatrix},
\]
each time extracting a vector from the nullspace to get our eigenvectors. Plugging \( x = 1 \) into this matrix is really unnecessary, because, we can already see from the standard matrix form that the first standard basis vector is an eigenvector, corresponding to the eigenvalue 1. Thus we will also use \((1,0,0)\) as the first basis vector in \(B\).

To get the second eigenvector, plug \( x = 2 \) into the matrix above to get

\[
\begin{pmatrix}
1 & -1 & -1 \\
0 & 0 & -2 \\
0 & 0 & -1
\end{pmatrix}.
\]

We see the dependency between the first and second columns, which leads us to immediately write down \((1,1,0)\) as a vector in the nullspace. This is then our second basis vector to put into \(B\). To get the last one, we plug \( x = 3 \) into the matrix above to get

\[
\begin{pmatrix}
2 & -1 & -1 \\
0 & 1 & -2 \\
0 & 0 & 0
\end{pmatrix}.
\]

We see that the last column is some combination of the first two columns; writing

\[
\alpha \begin{pmatrix} 2 \\ 0 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} -1 \\ -2 \\ 0 \end{pmatrix}
\]

and solving gives us \(\alpha = -\frac{3}{2}\) and \(\beta = -2\), and we use this to say that \((\frac{3}{2}, -2, -1)\) is in the nullspace of this matrix, and hence an eigenvector of \(T\). Thus, when \(B = \{(1,0,0), (1,1,0), (\frac{3}{2}, -2, -1)\}\), we get a diagonal form for \(T\), which is

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3
\end{pmatrix}_B.
\]

Problem 8 Assume \(T : C^3 \to C^3\) is the transformation with standard matrix

\[
\begin{pmatrix}
1 & 0 & 0 \\
2 & 2 & 0 \\
3 & 3 & 3
\end{pmatrix}.
\]

Find a basis \(B\) consisting of eigenvectors, and give the corresponding \(B\)-matrix.

Example 8 Find a basis \(B\) consisting of eigenvectors for the transformation \(T : C^3 \to C^3\) whose standard matrix is

\[
\begin{pmatrix}
1 & 1 & 1 \\
1 & 1 & 1 \\
1 & 1 & 1
\end{pmatrix}.
\]

Give the corresponding \(B\)-matrix.

The characteristic polynomial is the determinant of

\[
\begin{pmatrix}
x-1 & -1 & -1 \\
-1 & x-1 & -1 \\
-1 & -1 & x-1
\end{pmatrix},
\]

which is (expanding along the first row)

\[
(x-1)[(x-1)^2 - 1] - (-1)[(-1)(x-1) - 1] + (-1)[1 + (x-1)].
\]
After multiplying and grouping terms, this simplifies to \( x^3 - 3x^2 \) which factor as \( x^2(x - 3) \). From the factorization, we see that 3 is an eigenvalue without multiplicity, and 0 is an eigenvalue with multiplicity, which leads us to expect that the eigenspace corresponding to 0 is a plane, and the eigenspace corresponding to 3 is a line. Plug 3 into the matrix above to get

\[
\begin{pmatrix}
2 & -1 & -1 \\
-1 & 2 & -1 \\
-1 & -1 & 2
\end{pmatrix},
\]

for which we struggle to see a dependency between the columns;

\[
\alpha \begin{pmatrix} 2 \\ -1 \\ -1 \end{pmatrix} + \beta \begin{pmatrix} -1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ -1 \\ 2 \end{pmatrix}.
\]

The first two coordinates gives the system

\[
\begin{align*}
2\alpha - \beta &= -1 \\
-\alpha + 2\beta &= -1
\end{align*}
\]

which has the unique solution \( \alpha = -1 \) and \( \beta = -1 \). We then verify that equality also holds in the last coordinate: \((-1)(-1) + (-1)(-1) = 2 \). This dependency of columns tells us that \((-1, -1, -1)\) is in the nullspace of the above matrix, and hence this is an eigenvector of \( T \) corresponding to the eigenvalue 3.

We plug \( x = 0 \) into our matrix above, and get

\[
\begin{pmatrix}
-1 & -1 & -1 \\
-1 & -1 & -1 \\
-1 & -1 & -1
\end{pmatrix},
\]

which we see has rank 1, and hence a nullity of 2. This is expected, since we anticipated that the eigenspace corresponding to the eigenvalue 0 would be a plane. Thus, instead of extracting a single vector in the nullspace, we will take two independent ones. The dependency between the first and second columns tells us that \((1, -1, 0)\) is in the null space, and the dependency between the first and last columns tells us that \((1, 0, -1)\) is in the null space. These two vectors are independent and therefore span the two dimensional eigenspace of \( T \). Putting these eigenvectors into a basis, we arrive at \( \mathcal{B} = \{(-1, -1, -1), (1, -1, 0), (1, 0, -1)\} \), and the corresponding \( \mathcal{B} \)-matrix of \( T \) is

\[
\begin{pmatrix}
3 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}_B.
\]

**Problem 9** Find a basis \( \mathcal{B} \) consisting of eigenvectors for the transformation \( T : \mathbb{C}^3 \to \mathbb{C}^3 \) whose standard matrix is

\[
\begin{pmatrix}
1 & 1 & 0 \\
1 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}.
\]

Give the corresponding \( \mathcal{B} \)-matrix.

**Problem 10** Find a basis \( \mathcal{B} \) consisting of eigenvectors for the transformation \( T : \mathbb{C}^3 \to \mathbb{C}^3 \) whose standard matrix is

\[
\begin{pmatrix}
0 & 1 & 0 \\
-1 & 0 & 0 \\
0 & 0 & i
\end{pmatrix}.
\]
Chapter 4  •  Diagonalization

Give the corresponding $B$-matrix.

**Problem 11** Find a basis $B$ consisting of eigenvectors for the transformation $T : C^3 \to C^3$ whose standard matrix is

\[
\begin{pmatrix}
1 & 2 & 0 \\
0 & 2 & 1 \\
0 & 0 & 0
\end{pmatrix}.
\]

Give the corresponding $B$-matrix.

More Exercises
Chapter 5

Projections

5.1 Plane projections

Among the deepest and most central concepts of linear algebra is that of similarity. Two transformations are similar if there are two bases relative to which the corresponding matrices of the transformations are the same. Two similar transformations are essentially the manifestation of the same thing, except that they are viewed from different coordinate systems. We begin with the study of those transformations that are similar to the projections of a plane onto a coordinate axis.

Assume \( L \) denotes a line in \( \mathbb{R}^2 \) containing the zero vector; i.e., assume \( L \) is a one dimensional subspace of \( \mathbb{R}^2 \). The orthogonal projection of \( \mathbb{R}^2 \) onto \( L \) is the transformation that takes a vector in \( \mathbb{R}^2 \) to the vector on \( L \) closest to it. Notice that many vectors are taken to the same place: for example, every vector on the line through the origin and orthogonal to \( L \) is taken to the zero vector. Such a transformation is, consequently, not one-to-one, and not invertible.

A generalization of an orthogonal projection is obtained if we start with two lines in \( \mathbb{R}^2 \), \( L_1 \) and \( L_2 \), containing the zero vector. The projection of \( \mathbb{R}^2 \) onto \( L_1 \) in the direction of \( L_2 \) takes a vector \( \mathbf{v} \) to the unique vector \( T(\mathbf{v}) \) on the line \( L_1 \) with the line segment joining \( \mathbf{v} \) to \( T(\mathbf{v}) \) parallel to \( L_2 \). Thus the orthogonal projection of \( \mathbb{R}^2 \) onto \( L \) is the same as the projection onto \( L \) in the direction of the line perpendicular to \( L \).

There is a natural matrix form that describes every conceivable projection of \( \mathbb{R}^2 \) onto a line, and that is

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 
\end{pmatrix}
\]  

As the basis \( \mathcal{B} \) varies through all possibilities, this single matrix represents an infinite family of transformations. This family is exactly the set of all transformations similar to the orthogonal projection onto the x-axis. A natural matrix form, such as this, used to express a particular type of transformation, is called a canonical form.

Example 1

Assume \( T : \mathbb{R}^2 \rightarrow \mathbb{R}^2 \) is the orthogonal projection onto the line \( y = 7x \). Find vectors \( \mathbf{u} \) and \( \mathbf{v} \) so that the \( \mathcal{B} \)-matrix of \( T \) is

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 
\end{pmatrix}
\]

when \( \mathcal{B} = \{ \mathbf{u}, \mathbf{v} \} \), and use this to find the standard matrix of \( T \).

This projection has two nontrivial invariant subspaces: every vector on the line \( y = 7x \) is taken to itself, and every vector on the line orthogonal to \( y = 7x \); i.e., every vector on the line \( y = \frac{7}{20}x \), is taken to the zero vector. These two lines contain all of the eigenvectors of \( T \). Thus if we take \( \mathbf{u} \) on
the line $y = 7x$, for example, we let $u = (1, 7)$, and if we take $v$ on the line $y = \frac{-1}{7}x$, like $v = (7, -1)$, then $B = \{(1, 7), (7, -1)\}$ is a basis of eigenvectors, and the corresponding diagonal $B$-matrix is

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}
\]

To find the standard matrix, we need to see where $T$ sends $(1, 0)$ and $(0, 1)$, writing the result in the columns of the matrix using standard coordinates. We will do this by converting the vectors into $B$-coordinates, applying the $B$-matrix of $T$, then converting back to standard coordinates. Starting with the first basis vector, write

\[(1, 0) = \alpha(1, 7) + \beta(7, -1)\]

and solve, to get $\alpha = \frac{1}{50}$ and $\beta = \frac{7}{50}$, so the $B$-coordinates of $(1, 0)$ are $(\frac{1}{50}, \frac{7}{50})_B$. Now to see where $T$ sends this vector, we multiply

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
\frac{1}{50} \\
\frac{7}{50} \\
\end{pmatrix}
= \begin{pmatrix}
\frac{1}{50} \\
\frac{7}{50} \\
\end{pmatrix}
,
\]

then convert $(\frac{1}{50}, 0)_B$ back to standard coordinates by writing

\[
\frac{1}{50} \ (1, 7) + 0 \ (7, -1) = \left( \frac{1}{50}, \frac{7}{50} \right),
\]

giving the entries for the first column of the standard matrix; so far the standard matrix is

\[
\begin{pmatrix}
\frac{1}{50} & ? \\
\frac{7}{50} & ? \\
\end{pmatrix}
.
\]

Repeating the process for the second standard basis vector, we write

\[(0, 1) = \alpha(1, 7) + \beta(7, -1)\]

and solve, to get $\alpha = \frac{7}{50}$ and $\beta = \frac{1}{50}$, so the $B$-coordinates of $(0, 1)$ are $(\frac{7}{50}, \frac{1}{50})_B$. Now transform the vector

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}
\begin{pmatrix}
\frac{7}{50} \\
\frac{1}{50} \\
\end{pmatrix}
= \begin{pmatrix}
\frac{7}{50} \\
\frac{1}{50} \\
\end{pmatrix}
,
\]

and convert $(\frac{7}{50}, 0)$ back to standard coordinates

\[
\frac{7}{50} \ (1, 7) + 0 \ (7, -1) = \left( \frac{7}{50}, \frac{49}{50} \right),
\]

giving the entries for the second column, and our standard matrix is

\[
\begin{pmatrix}
\frac{1}{50} & \frac{7}{50} \\
\frac{7}{50} & \frac{49}{50} \\
\end{pmatrix}
.
\]

A quick visual check of the trace and the determinant reassures us that we have not made a mistake.

**Problem 1** Assume $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the orthogonal projection onto the line $y = 3x$. Find vectors $u$ and $v$ so that the $B$-matrix of $T$ is

\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}_B
\]

when $B = \{u, v\}$, and use this to find the standard matrix of $T$. 
Example 2
Assume $T$ is the projection onto the line $y = 3x$ in the direction of the line $y = x$. Find the standard matrix of $T$.

The invariant subspaces are the lines $y = 3x$ and $y = x$; every vector on the line $y = 3x$ is taken to itself, and every vector on the line $y = x$ gets transformed to the zero vector. It follows that $B = \{(1, 3), (1, 1)\}$ is a basis of eigenvectors, relative to which the $B$-matrix of $T$ is

\[
\begin{pmatrix}
1 & 0 \\
0 & 0
\end{pmatrix}_B.
\]

To get the standard matrix of $T$, begin by converting $(1, 0)$ into $B$-coordinates;

\[(1, 0) = \alpha(1, 3) + \beta(1, 1)\]

implies $\alpha = \frac{-1}{2}$, $\beta = \frac{3}{2}$, and $(1, 0) = (\frac{-1}{2}, \frac{3}{2})_B$. Transform the vector

\[
\begin{pmatrix}
1 & 0 \\
0 & 0
\end{pmatrix}_B \begin{pmatrix}
\frac{-1}{2} \\
\frac{3}{2}
\end{pmatrix}_B = \begin{pmatrix}
\frac{1}{2} \\
0
\end{pmatrix}_B,
\]

convert the result back to standard coordinates

\[-\frac{1}{2} (1, 3) + 0 (1, 1) = (\frac{-1}{2}, \frac{-3}{2}),\]

and we have the first column of the standard matrix.

Converting $(0, 1)$ into $B$-coordinates, we write

\[(0, 1) = \alpha(1, 3) + \beta(1, 1)\]

and solve, to get $\alpha = \frac{1}{2}$, $\beta = \frac{-1}{2}$, and $(0, 1) = (\frac{1}{2}, \frac{-1}{2})_B$. Transform the vector

\[
\begin{pmatrix}
1 & 0 \\
0 & 0
\end{pmatrix}_B \begin{pmatrix}
\frac{1}{2} \\
\frac{-1}{2}
\end{pmatrix}_B = \begin{pmatrix}
\frac{1}{2} \\
0
\end{pmatrix}_B,
\]

convert the result back to standard coordinates

\[\frac{1}{2} (1, 3) + 0 (1, 1) = (\frac{1}{2}, \frac{3}{2}),\]

and we have the second column, giving us the standard matrix

\[
\begin{pmatrix}
\frac{-1}{2} & \frac{1}{2} \\
\frac{-3}{2} & \frac{3}{2}
\end{pmatrix}.
\]

The trace is still 1 and the determinant is still 0, a very good omen.

Problem 2
Assume $T$ is the projection onto the line $y = 7x$ in the direction of the line $y = 3x$. Find the standard matrix of $T$.

When two different mathematical objects share identical structure, the term isomorphic is the adjective used. The meaning of isomorphic changes depending on the depth of structure within the objects; two isomorphic sets just have the same number of elements, isomorphic vector spaces have the same number of elements and they share identical vector addition and scalar multiplication, the space of all linear transformations on $R^3$ is isomorphic to the space of all their standard matrices.
because the underlying sets have the same number of elements, the addition and scalar multiplications correspond perfectly, and the composition corresponds perfectly to matrix multiplication. When an isomorphism is intuitively digested, an individual freely thinks of the one object as the other, and since the authors of linear algebra texts tend to think of matrices and linear transformations interchangeably, all of the terminology introduced for linear transformations also applies to their matrices. We have already begun to refer to the rank and nullity of a matrix, and we will also refer to two matrices being similar if they are the standard matrices of similar transformations.

There is another way to visualize which matrices are similar to what. Let $T$ denote the orthogonal projection onto the $x$-axis, and then imagine all of the $\mathcal{B}$-matrices of $T$

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

each one obtained as $\mathcal{B}$ varies through all possible bases. The set of matrices obtained in this way is exactly the set of all matrices similar to

$$
\begin{pmatrix}
  1 & 0 \\
  0 & 0 \\
\end{pmatrix}.
$$

The three invariants discussed in the previous chapter, trace, determinant, and characteristic polynomial, will be the same for every matrix in this group. The trace is always equal to 1, the determinant is 0, and the characteristic polynomial is always $x(x - 1)$. This group is called a similarity class; it consists of the set of all matrices similar to a given one.

**Example 3**

Determine which of the following is a standard matrix for a projection of $R^2$ onto a line:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
|i) | $\begin{pmatrix}
  \frac{1}{2} & -\frac{1}{2} \\
  -\frac{1}{2} & \frac{1}{2} \\
\end{pmatrix}$ | ii) $\begin{pmatrix}
  \frac{3}{2} & \frac{3}{2} \\
  -\frac{1}{2} & -\frac{1}{2} \\
\end{pmatrix}$ | iii) $\begin{pmatrix}
  \frac{3}{2} & -\frac{10}{12} \\
  \frac{3}{4} & \frac{1}{2} \\
\end{pmatrix}$ | iv) $\begin{pmatrix}
  0 & 0 \\
  -\frac{1}{2} & 1 \\
\end{pmatrix}$ | v) $\begin{pmatrix}
  2 & -\frac{1}{2} \\
  \frac{1}{2} & -1 \\
\end{pmatrix}$ |

We can eliminate the last one because one of the invariants gives the wrong answer; the determinant of such a projection must be 0, but the determinant of matrix (v) is not 0. It follows that matrix (v) is not a projection of $R^2$ onto a line.

For each of the other matrices, the traces are 1, the determinants are 0, and the characteristic polynomials are each $x(x - 1)$. If an invariant gives the wrong value, then we know that the matrix is not such a projection, but if an invariant gives the correct value, can we conclude that the matrix is the projection? The answer is no! An example is given in matrix (v), which gives the correct value for the trace even though the matrix is not a projection. A particularly good invariant that has the property that two matrices are similar exactly when the values of the invariant agree is called a complete similarity invariant, and the description of some general complete similarity invariants is our ultimate goal. We are still many pages away from this goal, but if we are willing to sacrifice generality, we are ready to meet a complete similarity invariant for projections of $R^2$ onto lines, and this is the characteristic polynomial. If the characteristic polynomial is $x(x - 1)$, then 1 and 0 are the eigenvalues, and when a corresponding basis $\mathcal{B}$ of eigenvectors is found, the resulting $\mathcal{B}$ matrix is

$$
\begin{pmatrix}
  1 & 0 \\
  0 & 0 \\
\end{pmatrix}_B,
$$

and that is what it means to be a projection of $R^2$ onto a line. It follows that the first four matrices represent projections of $R^2$ onto lines.

**Problem 3**

Determine which of the following is a standard matrix for a projection of $R^2$ onto a line:

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
|i) $\begin{pmatrix}
  3 & 3 \\
  -2 & -2 \\
\end{pmatrix}$ | ii) $\begin{pmatrix}
  -2 & -2 \\
  3 & 3 \\
\end{pmatrix}$ | iii) $\begin{pmatrix}
  3 & -2 \\
  -2 & 3 \\
\end{pmatrix}$ | iv) $\begin{pmatrix}
  -6 & 3 \\
  -2 & 1 \\
\end{pmatrix}$ |
Problem 4 Give an example of a matrix similar to
\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}
\]
whose northwest entry is 2. (Study the matrices in example 3 and problem 3 for inspiration.)

5.2 Space projections

Projections of three dimensional spaces come in two non-trivial flavors; there are the projections onto the one dimensional subspaces (lines), and the projections onto two dimensional subspaces (planes). We begin with the projections onto lines. Assume \( L \) is a line in \( \mathbb{R}^3 \) through the origin. The orthogonal projection onto \( L \) is the transformation that takes a vector \( \mathbf{u} \) to the unique element of \( L \) closest to it. Imagine the line \( L \) and imagine the plane through the origin orthogonal to \( L \). This plane is exactly the set of vectors that are taken to the zero vector on \( L \); i.e. this plane is the null space of the transformation. If \( \mathbf{u} \) is a vector not in the null space, we may visualize where \( \mathbf{u} \) is taken by imagining the plane parallel to the null space that contains \( \mathbf{u} \); then \( \mathbf{u} \) is taken to the intersection of this parallel plane with the line \( L \). In fact, every vector in this parallel plane is projected to the same element on the line \( L \).

Example 4

Assume \( T : \mathbb{R}^3 \rightarrow \mathbb{R}^3 \) is the orthogonal projection onto the line spanned by \( (1,3,2) \). Find a basis \( \mathcal{B} \) so that the \( \mathcal{B} \)-matrix of \( T \) is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}_{\mathcal{B}}.
\]
Use this to find the standard matrix of \( T \).

The eigenspace corresponding to the eigenvalue 1 is the line spanned by \( (1,3,2) \), and if we use any nonzero vector on this line as the first basis vector in \( \mathcal{B} \), the first column will be in the desired form. We might as well use the vector \( (1,3,2) \).

To find the second and third basis vectors, we need to use two independent vectors from the nullspace, which is the plane orthogonal to \( (1,3,2) \). To do this, we will use the ‘dot product’, which is covered in more detail later in the book. The dot product of \( (a,b,c) \) with \( (e,f,g) \) is defined to be \( ae + bf + cg \), and this equals zero exactly when \( (a,b,c) \) is orthogonal to \( (e,f,g) \); thus we seek two independent vectors \( (x,y,z) \) whose dot product with \( (1,3,2) \) is zero. Taking the dot product of \( (x,y,z) \) with \( (1,3,2) \), and setting this equal to zero, looks like
\[
x + 3y + 2z = 0.
\]

Grabbing the first solution that comes to mind, we see that \( (3,-1,0) \) is in the nullspace, and now we grab a second solution that is independent of \( (3,-1,0) \), like \( (2,0,-1) \). These two independent vectors span a plane, hence they span the nullspace of \( T \), and we will use them for the second and third basis vectors of \( \mathcal{B} \). If \( \mathcal{B} = \{(1,3,2),(3,-1,0),(2,0,-1)\} \), then the corresponding \( \mathcal{B} \)-matrix of \( T \) is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0 \\
\end{pmatrix}_{\mathcal{B}}.
\]

We now tackle the tedious chore of converting this matrix into standard coordinates. We need to perform the same sequence of steps three times, one time for each of the standard basis vectors. We convert the standard basis vector into \( \mathcal{B} \)-coordinates, use the \( \mathcal{B} \)-matrix to transform the vector,
then convert back to standard coordinates. We will show our work only for the first basis vector. Converting \((1, 0, 0)\) to \(B\)-coordinates, write
\[
(1, 0, 0) = \alpha(1, 3, 2) + \beta(3, -1, 0) + \gamma(2, 0, -1),
\]
and solve for \(\alpha\), \(\beta\), and \(\gamma\), getting \(\alpha = \frac{1}{14}\), \(\beta = \frac{3}{14}\), and \(\gamma = \frac{2}{14}\), so \((1, 0, 0) = (\frac{1}{14}, \frac{3}{14}, \frac{2}{14})_B\). Transform the vector with the \(B\)-matrix,
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}_B \begin{pmatrix}
\frac{1}{14} \\
\frac{3}{14} \\
\frac{2}{14}
\end{pmatrix}_B = \begin{pmatrix}
\frac{1}{14} \\
0 \\
0
\end{pmatrix}_B,
\]
and then convert the output \((\frac{1}{14}, 0, 0)_B\) back to standard coordinates
\[
\frac{1}{14}(1, 3, 2) + (0)(3, -1, 0) + (0)(2, 0, -1) = (\frac{1}{14}, \frac{3}{14}, \frac{2}{14}),
\]
which gives us the first column of our matrix. Repeating the process with the second and third standard basis vectors, we obtain the second and third columns of the standard matrix, giving us our answer
\[
\begin{pmatrix}
\frac{1}{14} & \frac{3}{14} & \frac{2}{14} \\
\frac{3}{14} & \frac{9}{14} & \frac{6}{14} \\
\frac{2}{14} & \frac{6}{14} & \frac{1}{14}
\end{pmatrix}_B.
\]
For those who enjoy appetizers, ponder the mysterious symmetry of the matrix across the main diagonal. One may also compare the answer just obtained with
\[
\begin{pmatrix}
1 \\
3 \\
2
\end{pmatrix}_B (1 3 2),
\]
which suggests an extremely fast method of obtaining a standard matrix for an orthogonal projection onto a line.

**Problem 5** Assume \(T : \mathbb{R}^3 \to \mathbb{R}^3\) is the orthogonal projection onto the line spanned by \((-1, 1, 2)\). Find a basis \(B\) so that the \(B\)-matrix of \(T\) is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}_B.
\]
Use this to find the standard matrix of \(T\).

There are also, of course, orthogonal projections onto planes. Assume that \(P\) is a plane in \(\mathbb{R}^3\) containing the zero vector. The orthogonal projection onto \(P\) is the transformation that takes a vector \(u\) to the unique element of \(P\) closest to it. Imagine the plane \(P\) and imagine the line \(L\) through the origin orthogonal to \(P\). The line \(L\) is exactly the set of vectors that are taken to the zero vector on \(P\), i.e., this line is the nullspace of the orthogonal projection onto \(P\). If \(u\) is a vector not in the null space, we may visualize where \(u\) is taken by imagining a line parallel to \(L\) that contains \(u\): then \(u\) is taken to the intersection of this parallel line with the plane \(P\), and the projection takes every vector on this line to the same place.

The orthogonal projections onto lines and planes are special cases of more general projections, just as they were in two dimensions. If \(L\) is a line through the origin, and \(P\) is a plane through the origin that does not contain \(L\) (\(P\) is not necessarily perpendicular to \(L\)), then we may define the projection onto \(L\) in the direction of \(P\) and the projection onto \(P\) in the direction of \(L\). The projection onto \(L\) takes all vectors on the plane \(P\) to the zero vector, and we visualize where a vector
not on \( P \) is taken by imagining the plane parallel to \( P \) containing \( \mathbf{u} \). Every vector on this parallel plane is taken to the intersection of \( L \) with this parallel plane. This projection onto \( L \) is thus a rank one transformation, and the orthogonal projection onto \( L \) is the special case when \( P \) is the plane perpendicular to \( L \). The projection onto \( P \) in the direction of \( L \) takes every vector on \( L \) to the zero vector, and vectors on lines parallel to \( L \) are all projected to the same place, namely the intersection of the parallel line with \( P \).

**Example 5**

Assume that \( P \) denotes the orthogonal projection onto the plane spanned by \((3, 1, 2)\) and \((2, 0, 1)\). Give a basis \( \mathcal{B} \) relative to which the \( \mathcal{B} \)-matrix of \( P \) is

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

Use this to find the standard matrix of \( P \).

Every nonzero vector on the plane spanned by \((3, 1, 2)\) and \((2, 0, 1)\) is an eigenvector corresponding to the eigenvalue 1, so to obtain the desired matrix form in the first two columns we need only select two independent vectors from this plane; let’s take \((3, 1, 2)\) and \((2, 0, 1)\). To obtain the desired column of zeros, the last basis vector should be taken from the nullspace. Those who know about cross products could use that knowledge now. Proceeding without any knowledge of cross products, we seek a vector \((x, y, z)\) that is orthogonal to both \((3, 1, 2)\) and \((2, 0, 1)\). When the dot product of \((x, y, z)\) is taken with \((3, 1, 2)\) we get zero, and likewise the dot product of \((x, y, z)\) with \((2, 0, 1)\) is zero; this statement gives us the two equations

\[
\begin{align*}
3x + y + 2z &= 0 \\
2x + z &= 0,
\end{align*}
\]

which, in matrix form, looks like

\[
\begin{pmatrix}
3 & 1 & 2 \\
2 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
x \\
y \\
z
\end{pmatrix}
= \begin{pmatrix}
0 \\
0
\end{pmatrix}.
\]

We can quickly solve this by seeing a dependency amongst the columns; twice the last column minus the middle column equals the first, which suggests a solution \((-1, -1, 2)\). It follows that this vector is in the nullspace of \( P \), and we will use it as the last vector in our basis \( \mathcal{B} \). Thus we take \( \mathcal{B} = \{(3, 1, 2), (2, 0, 1), (-1, -1, 2)\} \), and the corresponding \( \mathcal{B} \)-matrix of \( P \) is

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}
\]

To get the standard matrix of \( P \), we convert the standard basis vectors into \( \mathcal{B} \)-coordinates, use the \( \mathcal{B} \)-matrix to see how they are transformed, then convert the transformed vectors back to standard coordinates. To get the \( \mathcal{B} \)-coordinates of \((1, 0, 0)\), write

\[(1, 0, 0) = \alpha(3, 1, 2) + \beta(2, 0, 1) + \gamma(-1, -1, 2),\]

and solve for \( \alpha, \beta, \) and \( \gamma \), getting \( \alpha = -\frac{1}{5}, \beta = \frac{4}{5}, \) and \( \gamma = -\frac{1}{5}, \) so \((1, 0, 0) = (\frac{-1}{5}, \frac{4}{5}, -\frac{1}{5})_\mathcal{B}\). Transforming this vector, we have

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}
\begin{pmatrix}
\frac{-1}{5} \\
\frac{4}{5} \\
\frac{-1}{5}
\end{pmatrix}
= \begin{pmatrix}
\frac{-1}{5} \\
\frac{4}{5} \\
0
\end{pmatrix}.
\]
and converting \((-\frac{1}{6}, \frac{4}{6}, 0)_{B}\) to standard coordinates is accomplished by writing
\[
(-\frac{1}{6}, \frac{4}{6}, 0)_{B} = (-\frac{1}{6})(3, 1, 2) + (\frac{4}{6})(2, 0, 1) + (0)(-1, -1, 2) = (\frac{5}{6}, -\frac{1}{6}, \frac{2}{6}).
\]
So far, the standard matrix looks like
\[
\begin{pmatrix}
\frac{5}{6} & ? & ? \\
-\frac{1}{6} & ? & ? \\
\frac{2}{6} & ? & ?
\end{pmatrix}.
\]
Repeating the process for the second and third standard basis vectors, we end up with
\[
\begin{pmatrix}
\frac{5}{6} & -\frac{1}{6} & \frac{2}{6} \\
-\frac{1}{6} & \frac{5}{6} & \frac{2}{6} \\
\frac{2}{6} & \frac{2}{6} & \frac{2}{6}
\end{pmatrix}
\]
as the standard matrix for \(P\). The trace is still 2, so we are reassured that no mistake was made. Notice, once again, the mysterious symmetry across the main diagonal.

**Problem 6** Assume that \(P\) denotes the orthogonal projection onto the plane spanned by \((1, 1, 2)\) and \((1, 1, 1)\). Give a basis \(B\) relative to which the \(B\)-matrix of \(P\) is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}_{B}.
\]
Use this to find the standard matrix of \(P\).

**Example 6** Assume that \(P\) denotes the projection onto the plane spanned by \((3, 1, 2)\) and \((2, 0, 1)\) in the direction of the span of \((0, 1, 1)\). Give a basis \(B\) relative to which the \(B\)-matrix of \(P\) is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}_{B}.
\]
Use this to find the standard matrix of \(P\).

To visualize this projection onto a plane in the direction of a line, imagine holding the line in your hand and keeping your eye on the intersection of the line with the plane. Imaging that you can move the line, but not its orientation, so when it is moved it remains parallel to where it was. As the line is moved about, the coordinates of points that occupy the line change, and every conceivable coordinate can be made to lie on the line by moving it to an appropriate place. Wherever the line is in space, it represents exactly the set of points that project to the intersection with the plane. The span of \((0, 1, 1)\) determines the direction of the projection, and this line intersects the plane at the origin. So every point on this line is projected to the origin; i.e. this line is the nullspace. We can use any nonzero vector from this line as our third basis vector, and we will get our third column of zeros in the \(B\)-matrix of \(P\). We can use any two independent vectors from the plane as the first two basis vectors, so we will choose the most obvious ones, we take \((3, 1, 2)\) and \((2, 0, 1)\) again. Thus, our basis is \(B = \{(3, 1, 2), (2, 0, 1), (0, 1, 1)\}\) and the corresponding \(B\)-matrix is
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}_{B}.
\]
This time, we will only show the details of our work getting the second column of the standard matrix of $P$. To get the second column, we convert the second standard basis vector into $B$-coordinates, apply the transformation using its $B$-matrix, then convert this output back to standard coordinates. We write

$$(0, 1, 0) = \alpha(3, 1, 2) + \beta(2, 0, 1) + \gamma(0, 1, 1),$$

and solve, getting $\alpha = 2$, $\beta = -3$, and $\gamma = -1$, and $(0, 1, 0) = (2, -3, -1)_B$. We apply $P$, getting

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}_B
\begin{pmatrix}
2 \\
-3 \\
-1
\end{pmatrix}_B = \begin{pmatrix}
2 \\
-3 \\
0
\end{pmatrix}_B,
$$

and then transform the output $(2, -3, 0)_B$ into standard coordinates

$$(2, -3, 0)_B = (2)(3, 1, 2) + (-3)(2, 0, 1) + (0)(0, 1, 1) = (0, 2, 1).$$

Thus, the standard matrix looks something like

$$
\begin{pmatrix}
? & 0 & ? \\
? & 2 & ? \\
? & 1 & ?
\end{pmatrix}.
$$

**Problem 7** Assume that $P$ denotes the projection onto the plane spanned by $(1, 1, 2)$ and $(1, 1, 1)$ in the direction of the span of $(1, 2, 0)$. Give a basis $B$ relative to which the $B$-matrix of $P$ is

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 0
\end{pmatrix}_B.
$$

Use this to find the standard matrix of $P$.

### 5.3 An algebraic characterization of projections

The concept of a projection is a geometric one, though we have brought the tools of algebra with us by introducing a coordinate system relative to which the corresponding matrix of the projection is in diagonal form, with only 1’s and 0’s on the main diagonal. This is certainly an algebraic characterization of a projection, but not the one referred to in the title of this section. An even simpler, and more elegant, characterization is obtained as a corollary of the diagonal matrix form, and flushing out this characterization illustrates the true theme of this section; to see algebraic relationships between different transformations, put the transformations in their diagonal forms (if possible) where the relationships become selfevident.

**Example 7** Show that all of the projections $P$ we have discussed satisfy the identity $P^2 = P$.

Since $P$ is a projection, it has a diagonal form with 1’s and 0’s on the main diagonal. Let us write this matrix as

$$
\begin{pmatrix}
I & 0 \\
0 & 0
\end{pmatrix}_B,
$$

where it is understood that $I$ is an identity matrix of some size, and the 0’s are denoting zero matrices of various sizes. For example, the projection

$$
\begin{pmatrix}
1 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0
\end{pmatrix}_B,
$$
fits the form above when \( I \) is the \( 2 \times 2 \) identity matrix and the 0 in the lower right corner is denoting a \( 3 \times 3 \) zero matrix. The other two 0’s must then denotes a \( 2 \times 3 \) and a \( 3 \times 2 \). Such a matrix of matrices is often called a block matrix. Since this particular block matrix is diagonal, we have
\[
\begin{pmatrix}
1 & 0 \\
0 & 0
\end{pmatrix}^2_B = \begin{pmatrix}
I^2 & 0 \\
0 & 0^2
\end{pmatrix}_B = \begin{pmatrix}
I & 0 \\
0 & 0
\end{pmatrix}_B,
\]
so \( P^2 = P \).

**Problem 8** Assume \( P \) is a projection on some space, and \( I \) is the identity transformation on the same space. Show that \( I - P \) is also a projection.

We have established that any projection \( P \) satisfies \( P = P^2 \). If we are happy to consider the identity and zero transformations as projections, then we can establish that the converse is also true. More specifically, we can deduce that any transformation that satisfies \( P^2 = P \) is diagonalizable, and its eigenvalues are always either 0 or 1. Assume \( P^2 = P \) and \( \alpha \) is an eigenvalue for \( P \), so \( P(u) = \alpha u \) for a nonzero vector \( u \). Apply \( P \) to both sides to get \( P^2(u) = \alpha^2 u \), and use the fact that \( P^2 = P \) to conclude \( \alpha u = \alpha^2 u \). Since \( u \) is not zero, we must have \( \alpha = \alpha^2 \), and that means \( \alpha \) is either 1 or 0. The eigenspaces of 1 and 0 are disjoint, and the equation
\[
v = P(v) + (I - P)(v)
\]
is valid for every vector \( v \); but \( P(v) \) is in the eigenspace corresponding to 1 and \( (I - P)(v) \) is in the eigenspace corresponding to 0, so every vector lies in the span of these two eigenspaces. It follows that \( P \) is diagonalizable; if \( B \) is constructed by unioning a basis from each of these two eigenspaces, then \( B \) will be a basis of the domain, and the corresponding \( B \)-matrix of \( P \) is diagonal with 1’s and 0’s on the main diagonal.

**Example 8** Determine which of the following are the standard matrices of projections:

\[
\begin{align*}
i) & \quad \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & 1
\end{pmatrix} \\
ii) & \quad \begin{pmatrix}
0 & 0 & 0 \\
\pi & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} \\
iii) & \quad \begin{pmatrix}
0 & 0 & 0 \\
0 & 1 & \pi \\
0 & 0 & 1
\end{pmatrix} \\
iv) & \quad \begin{pmatrix}
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{1}{3} & \frac{1}{3}
\end{pmatrix}
\end{align*}
\]

With our algebraic characterization of a projection, all we need to do is square each matrix and see if it changed. The only one that is different than its square is the matrix in (iii); indeed,
\[
\begin{pmatrix}
0 & 0 & 0 \\
0 & 1 & \pi \\
0 & 0 & 1
\end{pmatrix}^2 = \begin{pmatrix}
0 & 0 & 0 \\
0 & 1 & 2\pi \\
0 & 0 & 1
\end{pmatrix}.
\]
Thus, this is the only one that is not a projection. Notice that the characteristic polynomial is \((x - 1)(x - 1)x\), just like that of any rank 2 projection on \( R^3 \). This means that in three dimensions, the characteristic polynomial is no longer a complete similarity invariant for the projections.

**Problem 9** Determine which of the following are the standard matrices of projections. For the ones that are projections, give a basis \( B \) of eigenvectors and write the corresponding \( B \)-matrix.

\[
\begin{align*}
i) & \quad \begin{pmatrix}
\frac{1}{2} & 0 & \frac{1}{2} \\
0 & 1 & 0 \\
\frac{1}{2} & 0 & \frac{1}{2}
\end{pmatrix} \\
ii) & \quad \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & 0 & 1
\end{pmatrix} \\
iii) & \quad \begin{pmatrix}
0 & \frac{1}{2} & 0 \\
\frac{1}{2} & 1 & \frac{1}{2} \\
0 & \frac{1}{2} & 0
\end{pmatrix} \\
iv) & \quad \begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 \\
0 & 1 & 0 \\
0 & 0 & \frac{1}{2}
\end{pmatrix}
\end{align*}
\]
Problem 10  Assume that $P : \mathbb{R}^3 \rightarrow \mathbb{R}^3$ denotes the projection onto a plane $M$ in the direction of a line $L$. What does the projection $I - P$ project onto, and in what direction?

More Exercises
Chapter 6

Reflections et cetera

6.1 Plane reflections

The projections are the diagonalizable transformations whose eigenvalues are 0 or 1. A **reflection** is a diagonalizable transformation whose eigenvalues are 1 or −1. If $T : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ is a reflection other than $I$ or $-I$, then there is a basis $B$ relative to which the matrix of $T$ is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_B,$$

which is a canonical form for reflections of the plane.

Just as for projections, a reflection of the plane is determined by two lines that intersect at the origin. In terms of a basis $B$ that puts the reflection in its canonical form, these two lines are the span of the two basis vectors in $B$. For the nonscalar reflections, each line is an eigenspace (an eigenline) corresponding to one of the two eigenvalues. An **orthogonal reflection** is one for which the two eigenlines are orthogonal. We use terminology that parallels the language introduced for projections; if $L_1$ is the span of $u_1$, $L_2$ is the span of $u_2$, $B = \{u_1, u_2\}$, and $T$ is the reflection with the $B$-matrix

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_B,$$

we say that $T$ is the reflection across $L_1$ in the direction of $L_2$. When $u_1$ is orthogonal to $u_2$, we refer to the reflection across $L_1$ in the direction of $L_2$ as the orthogonal reflection across $L_1$.

**Problem 1** Give a geometric description of the orthogonal reflection across the line spanned by $(1, 1)$, then give a basis $B$ relative to which the $B$-matrix of this orthogonal reflection is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_B.$$ 

**Problem 2** Give a geometric description of the reflection across the line spanned by $(1, 1)$ in the direction of $x$-axis, then give a basis $B$ relative to which the $B$-matrix of this reflection is

$$\begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}_B.$$ 

A canonical matrix form for a projection or a reflection lets us easily see, in an algebraic way, some relationships between the two transformations that are not so geometrically obvious.
Problem 3 Assume that $P : \mathbb{R}^2 \to \mathbb{R}^2$ is the projection onto $\mathcal{L}_1$ in the direction of $\mathcal{L}_2$. Use the canonical form for $P$ to show that $2P - I$ is the reflection across $\mathcal{L}_1$ in the direction of $\mathcal{L}_2$, then explain this relationship geometrically.

Problem 4 Assume that $T$ is the reflection of $\mathbb{R}^2$ across $\mathcal{L}_1$ in the direction of $\mathcal{L}_2$. Show that the projection of $\mathbb{R}^2$ onto $\mathcal{L}_1$ in the direction of $\mathcal{L}_2$ is the average of $T$ and the identity transformation.

6.2 Space reflections

Being able to clearly visualize a projection in $\mathbb{R}^3$ helps to see the effect of a space reflection. Assume that $P$ is the projection onto the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$, and imagine starting at a vector $v$ and travelling in a straight line toward $P(v)$. If you keep track of the distance travelled from $v$ to $P(v)$, then continue travelling past $P(v)$ an identical distance and in the same direction, you will end up at the reflection of $v$ across the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$. A vector that is on the plane $\mathcal{M}$ is thus taken to its reflection about the origin; i.e. every vector on the plane $\mathcal{M}$ is taken, by this reflection, to its negative. The reflection takes the vectors on $\mathcal{L}$ to themselves. Thus $\mathcal{M}$ is the two dimensional eigenspace, corresponding to the eigenvalue $-1$, and $\mathcal{L}$ is the one dimensional eigenspace of fixed points. If $u$ is a nonzero vector on the line $\mathcal{L}$, and $\{v, w\}$ is an independent subset of $\mathcal{M}$, then letting $B = \{u, v, w\}$ gives rise to the $B$-matrix

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & -1 & 0 \\
0 & 0 & -1
\end{pmatrix}
\]

for the reflection, which is its canonical form.

Problem 5 Assume that $P$ is the projection onto the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$, and $T$ is the reflection across the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$. Use canonical forms to show that $T = 2P - I$.

Problem 6 Assume that $P$ is the projection onto the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$, and $T$ is the reflection across the line $\mathcal{L}$ in the direction of the plane $\mathcal{M}$. Use canonical forms to show that $P - T$ is the projection onto the plane $\mathcal{M}$ in the direction of the line $\mathcal{L}$. Draw a picture that illustrates the geometry of this fact.

Projections are characterized by the simple algebraic relation $P^2 = P$, and reflections have an equally simple characterization, which is $T^2 = I$. To prove this assertion, one needs to show that $T^2 = I$ if and only if $T$ is diagonalizable with eigenvalues 1 and $-1$. Here is one direction of this argument: assume that $T^2 = I$. Then

\[
\left(\frac{T + I}{2}\right)^2 = \frac{T^2 + 2T + I}{4} = \frac{T + I}{2},
\]

so the average of $T$ and $I$ is a projection, by our characterization of projections. It follows that this average is diagonalizable with a block matrix of the form

\[
\begin{pmatrix}
I & 0 \\
0 & 0
\end{pmatrix}
\]

But

\[
T = 2 \left(\frac{T + I}{2}\right) - I,
\]
so the $B$-matrix of $T$ is
\[ 2 \begin{pmatrix} I & 0 \\ 0 & 0 \end{pmatrix}_B \begin{pmatrix} I \\ 0 \end{pmatrix} = \begin{pmatrix} I \\ 0 \end{pmatrix}_B = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}_B, \]
and we see from this matrix form that $T$ is diagonalizable with eigenvalues 1 and $-1$.

**Problem 7** Assume that $T$ is diagonalizable with eigenvalues 1 and $-1$. Use the canonical form for $T$ to show that $T^2 = I$.

### 6.3 Dilations

Perhaps the simplest linear transformation of all is the **scalar transformation**, which, due to its geometric action, is also called a **dilation**. The geometric effect of a dilation may be visualized in three dimensions by imagining a balloon being inflated, or, in two dimensions, by imagining the pupils of someone’s eyes immediately after they enter a dark room. A transformation that implements this geometric action is a scalar multiple of the identity transformation, so, in three dimensions, its standard matrix looks like
\[
\begin{pmatrix}
\alpha & 0 & 0 \\
0 & \alpha & 0 \\
0 & 0 & \alpha
\end{pmatrix}.
\]

Every nonzero vector is an eigenvector for this transformation, so the entire domain is an eigenspace.

**Problem 8** Assume that the standard matrix of $T : R^2 \to R^2$ is
\[
\begin{pmatrix}
\alpha & 0 \\
0 & \alpha
\end{pmatrix},
\]
and let $B = \{u, v\}$ be an arbitrary basis of $R^2$. Use the definition of a $B$-matrix to show that the $B$-matrix of $T$ is
\[
\begin{pmatrix}
\alpha & 0 \\
0 & \alpha
\end{pmatrix}_B.
\]

Problem 8 asserts that, if $T : R^2 \to R^2$ is a dilation, then there is only one matrix form for $T$. The converse of this statement is also true.

**Example 1** Assume that $T : R^2 \to R^2$ is a transformation for which every basis of $R^2$ yields the same matrix. Show that $T$ is a dilation.

Assume that $T$ is a transformation of $R^2$ for which every basis yields the same matrix. If the standard basis vectors are denoted $e_1$ and $e_2$, and the standard matrix of $T$ is
\[
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix},
\]
then with $B = \{e_2, e_1\}$, the $B$-matrix of $T$, which is
\[
\begin{pmatrix}
d & c \\
b & a
\end{pmatrix}_B,
\]
must be equal to the standard matrix, from which we deduce $a = d$ and $b = c$, and the standard matrix of $T$ is really
\[
\begin{pmatrix}
a & b \\
1 & a
\end{pmatrix}.
\]

We would now like to prove that every vector is an eigenvector for $T$, from which we will conclude that $b = 0$. Assume, by way of contradiction, that $u$ is not an eigenvector for $T$. Then $T(u)$ is not a multiple of $u$, so $B = \{u, T(u)\}$ is a basis of $R^2$ and the $B$ matrix of $T$ is of the form
\[
\begin{pmatrix}
0 & ? \\
1 & ?
\end{pmatrix}_B,
\]

which implies that the standard matrix is
\[
\begin{pmatrix}
0 & 1 \\
1 & 0
\end{pmatrix}.
\]

This is the matrix of the reflection across the line $y = x$, which does have a different matrix form (namely, a diagonal form with 1 and $-1$ on the main diagonal), which is contrary to our assumption that $T$ has only one matrix form. It follows that every vector is an eigenvector for $T$, and the standard matrix for $T$ must be
\[
\begin{pmatrix}
a & 0 \\
0 & a
\end{pmatrix},
\]

so $T$ is a dilation.

**Problem 9** Assume that $T : R^3 \to R^3$ is a transformation for which every basis of $R^3$ yields the same matrix. Show that $T$ is a dilation. (Hint: show that every vector is an eigenvector, as was done in the paragraph above.)

### 6.4 Shears and nilpotents

A shear is a transformation $T : R^2 \to R^2$ that has the matrix
\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}_B
\]
relative to some basis.

**Problem 10** Assume $T : R^2 \to R^2$ has the standard matrix
\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}.
\]

Describe the geometric effect of this transformation.

**Problem 11** Assume $T : R^2 \to R^2$ has the matrix
\[
\begin{pmatrix}
1 & a \\
0 & 1
\end{pmatrix}_B
\]
relative to a basis $B = \{u, v\}$. If $a \neq 0$, show that $T$ has a matrix form
\[
\begin{pmatrix}
1 & 1 \\
0 & 1
\end{pmatrix}_C
\]
relative to some basis $C$. 
Projections are characterized by the equation $P^2 = P$, and reflections by the equation $W^2 = I$. Will see that shears are characterized by the two equations $T \neq I$ and $(T - I)^2 = 0$. The hard direction to establish is, if $T$ satisfies these two equations, then show $T$ is a shear, i.e. show $T$ has a matrix of the form
\[
\begin{pmatrix}
1 & 1 \\
0 & 1 \\
\end{pmatrix}
\].

Try doing problem 12, and if a hint is needed, it can be found in the following paragraph.

**Problem 12** Show that $T : R^2 \to R^2$ is a shear if and only if $T \neq I$ and $(T - I)^2 = 0$.

A transformation $N$ is called a nilpotent if $N$ raised to some power is zero. Nilpotents are never diagonalizable, with one exception, and in some sense they serve as the prototype of nondiagonalizability, an idea that will be flushed out in more detail in a later chapter. The smallest power $m$ for which $N^m = 0$ is particularly important, because when we find a vector $u$ with $N^{m-1}u \neq 0$, we can show that the vectors
\[ u, Nu, \ldots, N^{m-1}u \]
are independent. When $m$ is the same as the dimension of the domain, these vectors will give us a basis, relative to which, the matrix of $N$ is particularly simple. This is the situation in problem 12, where $N = T - I$ and $m = 2$. The assumptions that $N \neq 0$ implies the existence of a vector $u \in R^2$ for which $Nu \neq 0$. One can then argue that $\{Nu, u\}$ forms a basis of $R^2$, and when we figure out what the corresponding matrix for $N$ looks like, we will know what the corresponding matrix of $T$ is as well.

**Example 2**

Assume $T : R^3 \to R^3$ is the nilpotent transformation whose standard matrix is
\[
\begin{pmatrix}
-3 & -2 & -1 \\
-6 & 2 & -8 \\
3 & 2 & 1 \\
\end{pmatrix}
\].

Find a vector $u$ for which $T^2u \neq 0$, and give the $B$-matrix of $T$ when $B = \{u, Tu, T^2u\}$.

We compute the standard matrix of $T^2$, to get
\[
\begin{pmatrix}
18 & 0 & 18 \\
-18 & 0 & -18 \\
-18 & 0 & -18 \\
\end{pmatrix}
\],

from which we see that either the first or last standard basis vector can serve for $u$; let us take $u = (0, 0, 1)$, so $T^2u = (18, -18, -18)$. Now if $B = \{u, Tu, T^2u\}$, then in this particular instance we have
\[
B = \left\{ \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}, \begin{pmatrix} -1 \\ -8 \\ 1 \end{pmatrix}, \begin{pmatrix} 18 \\ -18 \\ -18 \end{pmatrix} \right\},
\]
and the corresponding $B$-matrix of $T$ is
\[
\begin{pmatrix}
0 & 0 & 0 \\
1 & 0 & 0 \\
0 & 1 & 0
\end{pmatrix}
\].

**Problem 13** Assume $T : R^3 \to R^3$ is the transformation whose standard matrix is
\[
\begin{pmatrix}
-1 & 0 & 1 \\
-1 & -1 & 2 \\
-1 & 1 & 2
\end{pmatrix}
\].
Verify that $T^2 \neq 0$ but $T^3 = 0$, then find a vector $\mathbf{u}$ for which $T^2 \mathbf{u} \neq 0$, and give the $\mathcal{B}$-matrix of $T$ when $\mathcal{B} = \{ T^2 \mathbf{u}, T \mathbf{u}, \mathbf{u} \}$.

**Problem 14** Suppose $T : C^2 \to C^2$ has standard matrix

$$
\begin{pmatrix}
i & 1 \\
1 & -i
\end{pmatrix}.
$$

Find a basis $\mathcal{B}$ relative to which the $\mathcal{B}$-matrix of $T$ is

$$
\begin{pmatrix}
0 & 0 \\
1 & 0
\end{pmatrix}_B.
$$

### 6.5 Rotations

Assume $\theta$ is a real number, and consider the transformation of $R^2$ that rotates vectors counterclockwise by the angle $\theta$ (given in radian measure). Thus, the first standard basis vector $(1, 0)$ is taken to $(\cos \theta, \sin \theta)$, by the definition of the trigonometric functions. The second standard basis vector is $(0, 1)$, which is the same as $(\cos \frac{\pi}{2}, \sin \frac{\pi}{2})$. Our transformation takes this basis vector to $(\cos(\frac{\pi}{2} + \theta), \sin(\frac{\pi}{2} + \theta))$, which is the same as $(-\sin \theta, \cos \theta)$ (by the trig sum of angles formula). It follows that the standard matrix of the transformation is

$$
\begin{pmatrix}
\cos \theta & -\sin \theta \\
\sin \theta & \cos \theta
\end{pmatrix},
$$

which can be expressed without reference to the angle by writing

$$
\begin{pmatrix}
a & -b \\
b & a
\end{pmatrix},
$$

with the condition that $a^2 + b^2 = 1$.

The rotations described in the previous paragraph are transformations of the plane $R^2$, i.e. they are transformations with $T : R^2 \to R^2$, but their most powerful canonical form is obtained by considering them to be transformations of $C^2$, where these transformations are diagonalizable.

**Example 3**

Consider the rotation whose standard matrix is

$$
\begin{pmatrix}
\frac{\sqrt{2}}{2} & -\frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{pmatrix}.
$$

Find the canonical diagonal form for this transformation, and find the standard matrix of a different rotation that it is similar to.

We use the diagonalization routine to find the eigenvalues of this transformation. Since we are not asked to find the eigenvectors, we can stop as soon as we have the eigenvalues, since the diagonal form is obtained by writing a diagonal matrix with the eigenvalues on the main diagonal.

The characteristic polynomial is the determinant of

$$
\begin{pmatrix}
x - \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\
-\frac{\sqrt{2}}{2} & x - \frac{\sqrt{2}}{2}
\end{pmatrix},
$$
which, after a little algebra, and the use of the quadratic formula, is

\[(x - (\frac{\sqrt{2}}{2} + i\frac{\sqrt{2}}{2}))(x - (\frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2})).\]

It follows that the diagonal form for this transformation is

\[
\begin{pmatrix}
\frac{\sqrt{2}}{2} & i\frac{\sqrt{2}}{2} \\
0 & \frac{\sqrt{2}}{2} - i\frac{\sqrt{2}}{2}
\end{pmatrix}
\]

The rotation in this example is a special case of the general formula

\[
\begin{pmatrix}
a & -b \\
b & a
\end{pmatrix}
\]

when \(a = \frac{\sqrt{2}}{2}\) and \(b = \frac{\sqrt{2}}{2}\), and the angle of rotation is \(\theta = \frac{\pi}{4}\). When \(\theta = -\frac{\pi}{4}\), we obtain a standard matrix

\[
\begin{pmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\
-\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{pmatrix}
\]

which has the same characteristic polynomial, and hence the same diagonal form. It follows that the two rotations

\[
\begin{pmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\
-\frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2}
\end{pmatrix}
\]

are similar.

**Problem 15** Consider the rotation whose standard matrix is

\[
\begin{pmatrix}
\frac{\sqrt{3}}{2} & -\frac{1}{2} \\
\frac{1}{2} & \frac{\sqrt{3}}{2}
\end{pmatrix}
\]

Find the canonical diagonal form for this transformation, and find the standard matrix of a different rotation that it is similar to.

**Problem 16** Show that the matrices

\[
\begin{pmatrix}
a & -b \\
b & a
\end{pmatrix}
\]

and

\[
\begin{pmatrix}
a & b \\
-b & a
\end{pmatrix}
\]

are always similar. When \(a^2 + b^2 = 1\), show that one is the rotation by an angle \(\theta\) and the other by the angle \(-\theta\).

**Problem 17** Show that the roots of the characteristic polynomial for

\[
\begin{pmatrix}
a & -b \\
b & a
\end{pmatrix}
\]

are \(a + ib\) and \(a - ib\).

### 6.6 Block matrices

Many transformations of higher dimensional spaces are built up using building blocks that are themselves transformations of smaller spaces. Such a transformation is often described using a
**block matrix**, that is, a matrix whose entries are themselves matrices. If \( A \) and \( B \) denote the standard matrices of two transformations of \( \mathbb{R}^2 \) then
\[
\begin{pmatrix} A & B \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}
\]
are the standard matrices of two transformations of \( \mathbb{R}^4 \). The second of these is in **block diagonal form**, which means that the eigenvalues, characteristic polynomial, trace, and determinant can be easily related to that of \( A \) and \( B \); the characteristic polynomial of the block diagonal matrix is the product of the characteristic polynomials of \( A \) and \( B \), and the trace of the block diagonal matrix is the sum of the traces of \( A \) and \( B \). The first matrix is not a block diagonal matrix (unless \( B = 0 \)), but it is in block triangular form, so we still get useful information at a glance, just as we do for scalar triangular matrices.

The matrix algebra for block matrices works exactly as it does for scalar matrices, with the only exception that multiplication of the block entries is no longer commutative. Thus, we have
\[
\begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \begin{pmatrix} AE + BG & AF + BH \\ CE + DG & CF + DH \end{pmatrix},
\]
and in general \( AE + BG \neq EA + GB \).

**Problem 18** Describe the geometric effect of the following transformations:

(i) \[
\begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ 0 & \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix}
\]

(ii) \[
\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{pmatrix}
\]

(iii) \[
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix}
\]

**Example 4**

Find a basis \( B \) of eigenvectors for the transformation of \( \mathbb{R}^4 \) whose standard matrix is
\[
\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 1 & 0 \end{pmatrix},
\]
and give the corresponding \( B \)-matrix.

If \( A \) denotes the \( 2 \times 2 \) northwest block, and \( B \) denotes the \( 2 \times 2 \) southeast block, then we can find the eigenvectors of \( A \) and \( B \) and combine them to build a basis of eigenvectors for the \( 4 \times 4 \) matrix.

The matrix
\[
A = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}
\]
represents the orthogonal reflection across the \( x \)-axis, and consequently the standard basis vectors \((1,0)\) and \((0,1)\) are eigenvectors. This gives rise to two eigenvectors for the block matrix: when \( u \) is an eigenvector for \( A \) and \( o \) is the zero vector in \( \mathbb{R}^2 \), then \((u,o)\) is an eigenvector for the block matrix since
\[
\begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix} \begin{pmatrix} u \\ o \end{pmatrix} = \begin{pmatrix} Au \\ o \end{pmatrix}.
\]

Thus, we see that \((1,0,0,0)\) and \((0,1,0,0)\) are two eigenvectors of the block matrix. Similarly, the matrix
\[
B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}
\]
represents the orthogonal reflection across the line $y = x$, so $(1,1)$ and $(-1,1)$ are two eigenvectors of $B$, with corresponding eigenvectors $(0,0,1,1)$ and $(0,0,-1,1)$ for the block matrix. It follows that $\mathcal{B} = \{(1,0,0,0),(0,1,0,0),(0,0,1,1),(0,0,-1,1)\}$ is a basis of eigenvectors, relative to which the $\mathcal{B}$-matrix of the transformation is
\[
\begin{pmatrix}
1 & 0 & 0 & 0 \\
0 & -1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & -1
\end{pmatrix}_B.
\]

**Problem 19** Find a basis $\mathcal{B}$ of eigenvectors for the transformation of $\mathbb{R}^4$ whose standard matrix is
\[
\begin{pmatrix}
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
\frac{1}{2} & \frac{1}{2} & 0 & 0 \\
0 & 0 & -\frac{1}{2} & \frac{1}{2} \\
0 & 0 & \frac{1}{2} & -\frac{1}{2}
\end{pmatrix},
\]
and give the corresponding $\mathcal{B}$-matrix.

**Example 5**

Give the characteristic polynomial and trace of the transformation whose standard matrix is
\[
\begin{pmatrix}
1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 \\
0 & 0 & 2 & 2 \\
0 & 0 & 2 & 2
\end{pmatrix}.
\]

The characteristic polynomial of the northwest block is $(x-1)^2-1$, and the characteristic polynomial of the southeast block is $(x-2)^2-4$, so the characteristic polynomial for the block diagonal matrix is $[(x-1)^2-1][(x-2)^2-4]$. The trace is $1+1+2+2$.

**Problem 20** Give the characteristic polynomials and traces of the transformations whose standard matrices are
\[
(i) \quad \begin{pmatrix}
1 & 6 & 0 & 0 \\
0 & 3 & 0 & 0 \\
0 & 0 & 2 & 0 \\
0 & 0 & 5 & 4
\end{pmatrix}, \quad (ii) \quad \begin{pmatrix}
0 & -1 & 0 & 0 \\
1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 1 & 0
\end{pmatrix}, \quad (iii) \quad \begin{pmatrix}
1 & 0 & 6 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 3 & 0 \\
0 & 5 & 0 & 4
\end{pmatrix}
\]

**Problem 21** Show that every transformation $T : \mathbb{R}^3 \to \mathbb{R}^3$ has a $\mathcal{B}$-matrix of the form
\[
\begin{pmatrix}
a & b \\
0 & A
\end{pmatrix}_B,
\]
with $A$ being a $2 \times 2$ matrix. Discuss whether or not it is possible to obtain a block form with $b = 0$; i.e. can the basis $\mathcal{B}$ always be found so that the $\mathcal{B}$-matrix is of the block diagonal form
\[
\begin{pmatrix}
a & 0 \\
0 & A
\end{pmatrix}_B?
\]

More Exercises
Chapter 7

Generalization

By giving a broad interpretation to the concept of a vector, we will find that the theory of coordinates and linear transformations developed for \( \mathbb{R}^n \) and \( \mathbb{C}^n \) applies to many other spaces. The most prevalent generalizations apply to sets of real or complex valued functions, where the functions themselves are thought of as the vectors, with addition and scalar multiplication being the familiar pointwise operations defined by

\[
(f + g)(x) \equiv f(x) + g(x) \quad \text{and} \quad (\alpha f)(x) \equiv \alpha f(x).
\]

The definitions of independence, span, basis, and linear transformation carry over with no change.

7.1 When polynomials are the vectors

Let \( \mathcal{P}_n \) denote the set of polynomials of degree less than or equal to \( n \). When we need to distinguish between real polynomials and complex polynomials, we do so by writing \( \mathcal{P}_n(\mathbb{R}) \) and \( \mathcal{P}_n(\mathbb{C}) \), respectively. Thus,

\[
\mathcal{P}_n(\mathbb{R}) \equiv \{ a_0 + \ldots + a_n x^n : a_i \in \mathbb{R}, \ i = 0, \ldots, n \}
\]

and

\[
\mathcal{P}_n(\mathbb{C}) \equiv \{ a_0 + \ldots + a_n x^n : a_i \in \mathbb{C}, \ i = 0, \ldots, n \}.
\]

Expressing a polynomial in the customary way, by writing \( a_0 + \ldots + a_n x^n \), or to give specific examples, by writing \( 1 + 6x + 5x^2 \) or \( 9 - 2x^2 \), is actually expressing the polynomial in terms of a basis. The set \( \{ 1, x, \ldots, x^n \} \) is a basis of \( \mathcal{P}_n \), and relative to this basis the coordinates of the vector \( 1 + 6x + 5x^2 \) are plainly visible. If \( \sigma \) is a constant scalar, then \( \{ 1, x - \sigma, (x - \sigma)^2, \ldots, (x - \sigma)^n \} \) is a second basis of \( \mathcal{P}_n \), and writing a polynomial in terms of this basis is sometimes described by saying the polynomial is centered at \( \sigma \).

Example 1

Write the polynomial \( 1 + 6x + 5x^2 \) in terms of the basis \( \{ 1, (x-3), (x-3)^2 \} \), i.e. write the polynomial centered at 3.

We write

\[
1 + 6x + 5x^2 = \alpha 1 + \beta (x - 3) + \gamma (x - 3)^2;
\]

and solve for \( \alpha \), \( \beta \), and \( \gamma \). We multiply out the right hand side, getting

\[
1 + 6x + 5x^2 = (\alpha - 3\beta + 9\gamma) + (\beta - 6\gamma)x + \gamma x^2,
\]

then equate the three coefficients to obtain the system

\[
\begin{align*}
1 &= \alpha - 3\beta + 9\gamma \\
6 &= \beta - 6\gamma \\
5 &= \gamma,
\end{align*}
\]
which has the solution $\gamma = 5$, $\beta = 36$, and $\alpha = 64$, so
\[
1 + 6x + 5x^2 = 64 + 36(x - 3) + 5(x - 3)^2.
\]

**Problem 1** Write the polynomial $2 - x + 3x^2$ in terms of the basis $\{1, (x - 2), (x - 2)^2\}$, i.e. write the polynomial centered at 2.

In the calculus curriculum, students learn to relate a function $f$ with a Taylor series, which is a limit of polynomials. The coefficients of the Taylor series centered at $\sigma$ are given by the formula
\[
a_n = \frac{f^{(n)}(\sigma)}{n!},
\]
and when $f$ satisfies enough hypotheses, one obtains the equality
\[
f(x) = a_0 + a_1(x - \sigma) + a_2(x - \sigma)^2 + \ldots.
\]
When $f$ is itself a polynomial, the coefficients $a_n$ are exactly the ones needed to write $f$ centered at $\sigma$, so the formula that gives the Taylor coefficients may be thought of as a change of basis formula.

**Example 2** Use the Taylor coefficient formula to find the $B$-coordinates of $f(x) = (x - 3)(x^2 + 1)$ when $B = \{1, (x - 4), (x - 4)^2, (x - 4)^3\}$.

We are asked to find $a_1$, so that
\[
(x - 3)(x^2 + 1) = a_0 + a_1(x - 4) + a_2(x - 4)^2 + a_3(x - 4)^3,
\]
so in terms of the Taylor coefficient formula, we take $\sigma = 4$ and compute derivatives
\[
f^{(0)}(x) = (x - 3)(x^2 + 1), \quad f^{(1)}(x) = (x^2 + 1) + (x - 3)(2x), \quad f^{(2)}(x) = 2x + (2x) + 2(x - 3),
\]
and $f^{(3)}(x) = 6$. Now we plug in $\sigma$, getting
\[
f^{(0)}(4) = 17, \quad f^{(1)}(4) = 25, \quad f^{(2)}(4) = 18, \quad \text{and} \quad f^{(3)}(4) = 6.
\]
The Taylor coefficient formula now tells us that
\[
a_0 = \frac{17}{0!}, \quad a_1 = \frac{25}{1!}, \quad a_2 = \frac{18}{2!}, \quad \text{and} \quad a_3 = \frac{6}{3!},
\]
so $a_0 = 17$, $a_1 = 25$, $a_2 = 9$, and $a_3 = 1$. It follows that
\[
(x - 3)(x^2 + 1) = 17 + 25(x - 4) + 9(x - 4)^2 + (x - 4)^3,
\]
so the $B$-coordinates of $(x - 3)(x^2 + 1)$ are $(17, 25, 9, 1)_B$.

**Problem 2** Use the Taylor coefficient formula to find the $B$-coordinates of $f(x) = (x - 3)^2(x + 1)$ when $B = \{1, (x - 2), (x - 2)^2, (x - 2)^3\}$.

**Problem 3** Use the Taylor coefficient formula to find the $B$-coordinates of $f(x) = x^3$ relative to the basis $B = \{1, (x - i), (x - i)^2, (x - i)^3\}$.
of \( \mathcal{P}_3(C) \).

There are many transformations whose domain and codomain consist of polynomials, but perhaps instructors hesitate to tell their students that they are transformations the first time they are discussed. Two such transformations are the centerpiece of calculus, the differentiation and integration transformations. For polynomials, these transformations can be defined by writing

\[
D(a_0 + a_1 x + \ldots + a_n x^n) \equiv a_1 + 2a_2 x + 3a_3 x^2 + \ldots + na_n x^{n-1}
\]

and

\[
\int (a_0 + a_1 x + \ldots + a_n x^n) \equiv \int_0^x (a_0 + a_1 t + \ldots + a_n t^n) dt = a_0 x + \frac{a_1}{2} x^2 + \ldots + \frac{a_n}{n+1} x^{n+1}.
\]

Differentiation is a linear transformation because the derivative of a linear combination of functions is the same as the corresponding linear combination of the derivatives. The same thing can be said for integration.

**Example 3**

Let \( D : \mathcal{P}_2 \rightarrow \mathcal{P}_2 \) denote the differentiation transformation. Find the \( B \)-matrix of \( D \) when

\[
B = \{1, x^2 - x, x^2 + x\}.
\]

We need to compute \( D(1) \), \( D(x^2 - x) \), and \( D(x^2 + x) \), and write the results in the three columns of our \( B \)-matrix in terms of their \( B \)-coordinates. Computing derivatives, we find

\[
D(1) = 0, \quad D(x^2 - x) = 2x - 1, \quad \text{and} \quad D(x^2 + x) = 2x + 1.
\]

The \( B \)-coordinates of \( D(1) \) are \((0, 0, 0)_B\), so this is the first column of our matrix. The \( B \)-coordinates of \( D(x^2 - x) \) are obtained by writing

\[
2x - 1 = a_0 + a_1 (x^2 - x) + a_2 (x^2 + x),
\]

and solving for \( a_0 \), \( a_1 \), and \( a_2 \), from which we obtain \( a_0 = -1 \), \( a_1 = -1 \), and \( a_2 = 1 \), so \( D(x^2 - x) \) has \( B \)-coordinates \((-1, -1, 1)_B\), and we put these in the second column. So far, our \( B \)-matrix for \( D \) looks like

\[
\begin{pmatrix}
0 & -1 & \_ \\
0 & -1 & \_ \\
0 & 1 & \_ \\
\end{pmatrix}_B.
\]

Finally, we obtain \((1, -1, 1)_B\) as the \( B \)-coordinates of \( 2x + 1 \), so the \( B \)-matrix for \( D \) is

\[
\begin{pmatrix}
0 & -1 & 1 \\
0 & -1 & -1 \\
0 & 1 & 1 \\
\end{pmatrix}_B.
\]

**Problem 4**

Let \( D : \mathcal{P}_2 \rightarrow \mathcal{P}_2 \) denote the differentiation transformation. Find the \( B \)-matrix of \( D \) when

\[
B = \{1, x - 1, (x - 1)^2\}.
\]

We usually think of polynomials as functions, so when we use the symbol “1” to denote a polynomial, we mean that 1 is denoting the constant function whose graph is the horizontal line \( y = 1 \). Similarly, when we use the symbol \( x \) to denote a polynomial, we mean the function whose graph is the line \( y = x \). A more careful exposition would make an effort to distinguish between the
polynomial as a function, and the inputs and outputs of the function; for example we could define $u_i$ to be the polynomial given by the formula $u_i(x) = x^i$ (for $i = 0, 1, \ldots, n$). Now we can write \{u_0, u_1, u_2\} instead of \{1, x, x^2\}, which is really what we mean, because the former represents three polynomials while the latter represents three arbitrary outputs of these polynomials. It is, however, common in mathematical writing to see $u_2$ referred to as $x^2$, so it becomes the student’s responsibility to use the context of a symbol’s usage to decide whether the symbol represents a function or an output. Even though the symbol “1” means a real number at times, and it means a constant function at other times, it is the context of usage that tells us which of the meanings apply. If we wish to refer to an arbitrary element of $P_n$, we might symbolize it by writing $a_0 + a_1 x + \ldots + a_n x^n$ or we might simply write $p$; the second symbol is more accurate in that it actually denotes the function, while the first symbol is frequently used even though it is technically denoting $p(x)$, which is not denoting a function, but rather an output of the function.

Example 4

Assume $T : P_2 \rightarrow P_2$ is the transformation defined by

$$T(p) = p(1) + p(0)x$$

for every $p \in P_2$. Find a basis of eigenvectors for $T$ and give the corresponding matrix for $T$.

Let us first find any matrix for $T$; let $B = \{u_0, u_1, u_2\}$, where $u_0(x) = 1$, $u_1(x) = x$, and $u_2(x) = x^2$. We compute $T(u_i)$ and write the result in the $i^{th}$ column of the $B$-matrix, using the $B$-coordinates. By the definitions of $T$ and $u_0$, we have

$$T(u_0) = u_0(1) + u_0(0)x = 1 + x = (1, 1, 0)_B,$$

so the $B$-matrix of $T$ looks like

$$ \left( \begin{array}{ccc} 1 & ? & ? \\ 1 & ? & ? \\ 0 & ? & ? \end{array} \right)_B $$

so far. We repeat the process with $u_1$ and $u_2$, getting

$$T(u_1) = u_1(1) + u_1(0)x = 1 = (1, 0, 0)_B,$$

and

$$T(u_2) = u_2(1) + u_2(0)x = 1 = (1, 0, 0)_B,$$

so the $B$-matrix of $T$ is

$$ \left( \begin{array}{ccc} 1 & 1 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{array} \right)_B $$

To find the eigenvalues of $T$, we compute the determinant of

$$ \left( \begin{array}{ccc} x - 1 & -1 & -1 \\ -1 & x & 0 \\ 0 & 0 & x \end{array} \right)_B,$$

then find the roots of the resulting characteristic polynomial. This determinant is

$$x^3 - x^2 - x,$$

which is the same as $x(x^2 - x - 1)$, and we see that the roots are $x = 0$, $x = \frac{1 + \sqrt{5}}{2}$, and $x = \frac{1 - \sqrt{5}}{2}$.
7.2  •  When the vectors are matrices  77

Plugging $x = 0$ into the matrix above, we have

$$
\begin{pmatrix}
-1 & -1 & -1 \\
-1 & 0 & 0 \\
0 & 0 & 0
\end{pmatrix}_B,
$$

and the dependency of the last two columns suggests that $(0, 1, -1)_B$ is a vector in the nullspace. It follows that

$$(0, 1, -1)_B = x - x^2$$

is an eigenvector corresponding to the eigenvalue $0$. We repeat this process with the other two eigenvalues, beginning with $x = \frac{1+\sqrt{5}}{2}$, obtaining the matrix

$$
\begin{pmatrix}
\frac{1+\sqrt{5}}{2} - 1 & -1 & -1 \\
-1 & \frac{1+\sqrt{5}}{2} & 0 \\
0 & 0 & \frac{1+\sqrt{5}}{2}
\end{pmatrix}_B = \begin{pmatrix}
\frac{\sqrt{5}-1}{2} & -1 & -1 \\
-1 & \frac{\sqrt{5}+1}{2} & 0 \\
0 & 0 & \frac{1+\sqrt{5}}{2}
\end{pmatrix}_B,
$$

and looking for a dependency amongst the columns, we see that the second column is a multiple of the first, more specifically

$$
-2 \frac{\sqrt{5}-1}{\sqrt{5}-1} = \begin{pmatrix}
\frac{\sqrt{5}-1}{2} \\
-1 \\
0
\end{pmatrix}_B = \begin{pmatrix}
-1 \\
\frac{\sqrt{5}+1}{2} \\
0
\end{pmatrix}_B.
$$

It follows that

$$(\frac{-2}{\sqrt{5} - 1}, 1, 0)_B = \frac{-2}{\sqrt{5} - 1} - x$$

is an eigenvector corresponding to the eigenvalue $\frac{1+\sqrt{5}}{2}$. Similarly, we find that

$$(\frac{2}{\sqrt{5} + 1}, 1, 0)_B = \frac{2}{\sqrt{5} + 1} - x$$

is an eigenvector corresponding to the eigenvalue $\frac{1-\sqrt{5}}{2}$, so if

$$
C = \{x - x^2, \frac{-2}{\sqrt{5} - 1} - x, \frac{2}{\sqrt{5} + 1} - x\},
$$

then $C$ is a basis of eigenvectors, relative to which the $C$-matrix of $T$ is

$$
\begin{pmatrix}
0 & 0 & 0 \\
0 & \frac{1+\sqrt{5}}{2} & 0 \\
0 & 0 & \frac{1-\sqrt{5}}{2}
\end{pmatrix}_C.
$$

**Problem 5** Assume $T : \mathcal{P}_2 \rightarrow \mathcal{P}_2$ is the transformation defined by

$$
T(p) = p(1) + p(-1)x
$$

for every $p \in \mathcal{P}_2$. Find a basis of eigenvectors for $T$ and give the corresponding matrix for $T$.

**Problem 6** Assume $D : \mathcal{P}_2 \rightarrow \mathcal{P}_1$ is the differentiation transformation, and $\int : \mathcal{P}_1 \rightarrow \mathcal{P}_2$ is the integration transformation. Find a matrix representation of $\int \circ D : \mathcal{P}_2 \rightarrow \mathcal{P}_2$, and use it to show that $\int \circ D$ is a rank 2 projection.

**Problem 7** Assume $\int : \mathcal{P}_2 \rightarrow \mathcal{P}_3$ is the integration transformation, and $D : \mathcal{P}_3 \rightarrow \mathcal{P}_2$ is the differentiation transformation. What transformation is $D \circ \int$?
7.2 When the vectors are matrices

Let $M_{nm}$ denote the set of all $n \times m$ matrices ($n$ rows and $m$ columns), and when we need to emphasize which scalars are used, we will do so by writing $M_{nm}(R)$ or $M_{nm}(C)$. When $m = n$, we write $M_n$ to denote the space of $n \times n$ matrices. If $A \in M_n$, then each element of $M_{nm}$ can be multiplied on the left by $A$, which induces a transformation $L_A$ that is called a multiplication transformation. The function $L_A : M_{nm} \to M_{nm}$ is defined by

$$L_A(B) \equiv AB,$$

and the distributivity of matrix multiplication shows that $L_A$ is linear:

$$L_A(\alpha_1B_1 + \alpha_2B_2) = A(\alpha_1B_1 + \alpha_2B_2) = \alpha_1AB_1 + \alpha_2AB_2 = \alpha_1L_A(B_1) + \alpha_2L_A(B_2).$$

Similarly, when $A \in M_n$ there is a function $R_A : M_{nn} \to M_{nn}$ that multiplies on the right by $A$, and this is also a linear transformation for the same reason that $L_A$ is.

**Example 5**

Assume that $A \in M_2$ and $R_A : M_2 \to M_2$. If

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

and

$$B = \{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \}$$

is a basis of $M_2$, find the $B$-matrix of $R_A$.

We need to apply the transformation to the vectors in the basis $B$, then write the results in a matrix in terms of the output’s $B$-coordinates. Let us write $E_1$, $E_2$, $E_3$, and $E_4$, for the four vectors in $B$, and beginning with the first basis vector, we compute

$$R_A(E_1) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} = \begin{pmatrix} a & b \\ 0 & 0 \end{pmatrix} = (a,0,b,0)_B,$$

and we have the first column of our $B$-matrix, which looks like

$$\begin{pmatrix} a & 0 & 0 & 0 \\ 0 & a & 0 & 0 \\ b & 0 & d & 0 \\ 0 & b & 0 & d \end{pmatrix}_B$$

so far. We continue this process, obtaining $R_A(E_2) = (0,a,0,b)_B$, $R_A(E_3) = (c,0,d,0)_B$, and $R_A(E_2) = (0,c,0,d)_B$, so the $B$-matrix of $R_A$ is

$$\begin{pmatrix} a & 0 & c & 0 \\ 0 & a & 0 & c \\ b & 0 & d & 0 \\ 0 & b & 0 & d \end{pmatrix}_B.$$

**Problem 8**

Assume that $A \in M_2$ and $L_A : M_2 \to M_2$. If

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$$
and
\[ B = \{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \} \]
is a basis of \( M_2 \), find the \( B \)-matrix of \( L_A \).

**Problem 9** Assume that \( A \in M_2 \) and \( R_A : M_2 \rightarrow M_2 \). If
\[ A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \]
and
\[ B = \{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix} \} \]
is a basis of \( M_2 \), find the \( B \)-matrix of \( R_A \).

**Example 6**

Assume
\[ A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \]
Find the null space of \( L_A \).

If we write
\[ B = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \]
then we see that \( L_A(B) \) is
\[ \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \]

One way to express the answer is to go ahead and multiply the matrices out and set the result equal to zero;
\[ \begin{pmatrix} a + c & b + d \\ a + c & b + d \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}. \]

Thus \( B \) is in the nullspace of \( L_A \) if and only if \( a = -c \) and \( b = -d \), so the nullspace is exactly the set of matrices of the form
\[ \begin{pmatrix} -a & b \\ -a & -b \end{pmatrix}. \]

**Problem 10** Assume
\[ A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \]
Find the null space of \( R_A \).

**Example 7** Assume \( A \in M_n \) and \( \alpha \) is an eigenvalue of \( L_A : M_{nm} \rightarrow M_{nm} \). Show that \( \alpha \) is an eigenvalue of \( A \).

Assume \( A \in M_n \) and \( \alpha \) is an eigenvalue of \( L_A \). It follows that there is a nonzero matrix \( B \in M_{nm} \) with
\[ L_A(B) = \alpha B. \]

This equation is the same as \( AB = \alpha B \), and by the definition of matrix multiplication this means
\[ Au = \alpha u \]
for every column vector \( u \) in the matrix \( B \). We conclude that every column vector of \( B \)
is either the zero vector or an eigenvector for $A$; since $B$ is assumed nonzero, we see that one such

$u$ must be an eigenvector for $A$ corresponding to $\alpha$, and $\alpha$ is thus an eigenvalue for $A$.

**Problem 11** Assume $A \in M_2$ and $\alpha$ is an eigenvalue of $A$. Show that $\alpha$ is an eigenvalue of $L_A : M_{nn} \rightarrow M_{nn}$.

**Problem 12** Show that the eigenvalues of $A \in M_2$ are the same as the eigenvalues of $R_A : M_2 \rightarrow M_2$.

**Example 8**

Find the eigenvalues and eigenvectors of $L_A : M_2 \rightarrow M_2$ when

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$  

The eigenvalues of $L_A$ are the same as the eigenvalues of $A$ (by example 7 and problem 11), which are visible on the diagonal of the triangular form for $A$. The triangular form shows that the first

standard basis vector $(1, 0)$ is an eigenvector for $A$ corresponding to the eigenvalue 1. When we construct a matrix by putting this eigenvector in one column, and the zero vector in the other column, we obtain an eigenvector for $L_A$ corresponding to the same eigenvalue. There are two ways to do this, giving us the two matrices

$$\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$  

This gives us two independent eigenvectors for $L_A$. We obtain two more eigenvectors for $L_A$ by repeating this process with the other eigenvalue of $A$. We find an eigenvector of $A$ corresponding to the eigenvalue 2, which leads us to $(1, 1)$, then place this vector in one column, and the zero vector in the other column, giving us

$$\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}.$$  

Now we have that

$$B = \{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix} \}$$

is a basis of $M_2$, relative to which, the $B$-matrix of $L_A$ is

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{pmatrix}_B.$$

**Problem 13** Find the eigenvalues and a basis $B$ of eigenvectors for $L_A$ when

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 3 \end{pmatrix},$$

and give the corresponding $B$-matrix.

### 7.3 When the vectors are solutions to D.E.'s

An important application of linear transformations occurs in the study of differential equations. The simplest differential equations are first encountered in a first course of Calculus, where students are asked to find anti-derivatives of functions. One might phrase such a problem by asking for all solutions of the equation

$$f'(x) = x^2;$$
the answer is \( f(x) = \frac{x^3}{x} + C \). A differential equation has functions as its variables, and the equation expresses relationships between the function and its higher derivatives. Such equations arise naturally in applications; an analysis of a car bouncing on its springs leads one to an equation like

\[ f''(t) = -f(t). \]

A solution is a function \( f \) that satisfies this equation and it represents the vertical displacement of the car from its equilibrium position (resting on its springs). The connection with linear transformations occurs from the realization that any linear combination of solutions to such an equation is also a solution, so the set of solutions forms a vector space. One can then use the methods of linear transformations to analyze the solutions. Notice that \( \sin t \) and \( \cos t \) are solutions to the equation \( f''(t) = -f(t) \); as it happens, the set \( \mathcal{B} = \{\sin t, \cos t\} \) is a basis of the solution space for this equation.

When looked at from a certain perspective, it becomes obvious that solutions to certain differential equations form a vectors space. If we keep in mind that the operation of taking a derivative is a linear transformation, then we might emphasize this transformation viewpoint by writing the two equations above as \( D(f) = g \) and \( D^2f = -f \), where \( g(x) = x^2 \). The solutions to the equation \( D(f) = g \) do not form a vector space (the zero function is not a solution), but the solutions to \( D^2f = -f \) consist exactly of the null space of \( D^2 + 1 \), and that is why the solutions constitute a vector space. The study of linear differential equations essentially involves determining the null spaces of differential operators of the form

\[ a_0 + a_1 D + a_2 D^2 + \ldots + a_n D^n. \]

### 7.4 Transformations from one space to another

If \( T : \mathcal{V} \to \mathcal{W} \) satisfies \( T(\alpha \mathbf{u} + \beta \mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}) \) for every linear combination in \( \mathcal{V} \), then a basis of \( \mathcal{V} \) and a basis of \( \mathcal{W} \) will allow us to write a matrix for \( T \) that generalizes the idea of a \( \mathcal{B} \)-matrix when \( \mathcal{V} = \mathcal{W} \). The difference is that we now need two bases, a basis for the domain, that we will call \( \mathcal{B} \), and a basis for the codomain, that we will call \( \mathcal{C} \). To get the matrix for \( T \) relative to these two bases, we apply \( T \) to the basis vectors of \( \mathcal{B} \), then write these outputs in the columns of the matrix, using their \( \mathcal{C} \)-coordinates this time. The resulting matrix is called the \( \mathcal{B} \mathcal{C} \)-matrix for \( T \). The symbols representing the bases of the domain and the codomain are determined by their positions: if we speak of a \( \mathcal{B} \mathcal{C} \)-matrix of \( T \), we then intend \( \mathcal{C} \) to represent a basis of the domain, and \( \mathcal{B} \) now represents a basis of the codomain.

**Example 9**

Find the \( \mathcal{B} \mathcal{C} \)-matrix of the transformation \( T : M_2 \to \mathcal{P}_3 \) defined by

\[
T(\begin{pmatrix} a & b \\ c & d \end{pmatrix}) = a + bx + cx^2 + dx^3,
\]

when

\[
\mathcal{B} = \{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \}.
\]

and \( \mathcal{C} = \{ 1, (x-1), (x-1)^2, (x-1)^3 \} \).

We will show how to find the third column of the \( \mathcal{B} \mathcal{C} \)-matrix: apply \( T \) to the third basis vector and then determine the \( \mathcal{C} \)-coordinates of the result. We have

\[
T(\begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}) = x^2,
\]
and to find the $C$-coordinates write

$$x^2 = a_0 + a_1(x-1) + a_2(x-1)^2 + a_3(x-1)^3$$

and solve for $a_0$, $a_1$, $a_2$, and $a_3$. We will use the formula

$$a_i = \frac{f^{(i)}(1)}{i!},$$

with $f(x) = x^2$, getting $a_0 = 1$, $a_1 = 2$, $a_2 = 1$, and $a_3 = 0$, so

$$x^2 = (1, 2, 1, 0)_C,$$

and the $B,C$-matrix of $T$ looks something like

$$\begin{pmatrix}
? & ? & 1 & ? \\
? & ? & 2 & ? \\
? & ? & 1 & ? \\
? & ? & 0 & ? \\
\end{pmatrix}_{B,C}.$$

Repeating this process with the first, second, and fourth basis vectors, we obtain the $B,C$-matrix

$$\begin{pmatrix}
1 & 1 & 1 & 1 \\
0 & 1 & 2 & 3 \\
0 & 0 & 1 & 3 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}_{B,C}.$$

This matrix allows us to compute the action of $T$ in terms of the bases $B$ and $C$. For example, the matrix

$$\begin{pmatrix}
3 & 2 \\
1 & 0 \\
\end{pmatrix}$$

has $B$-coordinates $(3, 2, 1, 0)_B$, and when multiplied on the right of our $B,C$-matrix we obtain

$$\begin{pmatrix}
1 & 1 & 1 & 1 \\
0 & 1 & 2 & 3 \\
0 & 0 & 1 & 3 \\
0 & 0 & 0 & 1 \\
\end{pmatrix}_{B,C} \begin{pmatrix}
3 \\
2 \\
1 \\
0 \\
\end{pmatrix}_B = \begin{pmatrix}
6 \\
4 \\
1 \\
0 \\
\end{pmatrix}_C,$$

which are the $C$ coordinates of the output of $T$,

$$(6, 4, 1, 0)_C = 6 + 4(x-1) + (x-1)^2.$$ 

If no mistake was made, this should be the same as the output given by the definition of $T$, which is $3 + 2x + x^2$.

**Problem 14** Assume $D : P_3 \to P_2$ is the differentiation transformation. Find the $B,C$-matrix of $D$ when

$$B = \{1, x, x^2, x^3\} \quad \text{and} \quad C = \{1, (x+1), (x+1)^2\}.$$ 

**Problem 15** Assume $\int : P_2 \to P_3$ is the integration transformation. Find the $C,B$-matrix of $\int$ when

$$C = \{1, (x+1), (x+1)^2\} \quad \text{and} \quad B = \{1, x, x^2, x^3\}.$$
7.5 Isomorphisms and change of basis

There is a functional perspective of coordinates that begins with the observation that a given basis $B = \{e_1, \ldots, e_n\}$ of a real (or complex) space gives rise to an isomorphism $\iota_B$ of that space with $R^n$ (or $C^n$), the isomorphism being defined by

$$\iota_B(\alpha_1, \ldots, \alpha_n) = \alpha_1 e_1 + \ldots + \alpha_n e_n.$$  

The independence of $B$ corresponds to the fact that $\iota_B$ is one-to-one, and $\iota_B$ is onto precisely because $B$ is a spanning set. Notice that the vector $(\alpha_1, \ldots, \alpha_n)$ is taken to the vector that we have been denoting $(\alpha_1, \ldots, \alpha_n)_B$; in essence, the theory of coordinates that we developed is being supplanted by the study of isomorphisms.

Assume that $V$ is a vector space with a basis $B = \{e_1, \ldots, e_n\}$, and assume that $T : V \to V$ is a linear transformation. We then get a transformation $\iota_B^{-1} \circ T \circ \iota_B : R^n \to R^n$, which may be visualized using a diagram like

$$
\begin{array}{c}
R^n \\
\downarrow \iota_B^{-1}
\end{array}
\begin{array}{c}
\uparrow \iota_B \\
V
\end{array}
\begin{array}{c}
\downarrow T
\end{array}
\begin{array}{c}
V
\end{array}
\begin{array}{c}
\uparrow \iota_B^{-1}
\end{array}
\begin{array}{c}
R^n
\end{array}
$$

where the topmost right arrow represents the composition $\iota_B^{-1} \circ T \circ \iota_B$. A diagram, such as this one, is called a commutative diagram when it expresses the equality of composition of functions. The standard matrix of $\iota_B^{-1} \circ T \circ \iota_B$ has exactly the same entries as the $B$-matrix for $T$, and just as the isomorphism $\iota_B$ supplants the concept of $B$-coordinates, the standard matrix of $\iota_B^{-1} \circ T \circ \iota_B$ supplants the concept of a $B$-matrix.

**Example 10**

Let $D$ denote the differentiation transformation, assume $f(x) = x$, and define $T : \mathcal{P}_2 \to \mathcal{P}_2$ by

$$T(p) = D(fp).$$

If $B = \{1, x, x^2\}$, verify that the $B$-matrix of $T$ has the same entries as the standard matrix of $\iota_B^{-1} \circ T \circ \iota_B$ by computing them both.

Let us start by computing the $B$-matrix of $T$, beginning with the first column. Let $p$ be the first basis vector of $B$, so $p$ is the constant 1 function. It follows that

$$T(p) = D(fp) = D(f) = (1, 0, 0)_B,$$

which gives us the first column of the $B$-matrix. When $p(x) = x$, we have

$$T(p) = D(fp) = D(x^2) = (0, 2, 0)_B,$$

and when $p(x) = x^2$, we have

$$T(p) = D(fp) = D(x^3) = (0, 0, 3)_B,$$

so the $B$-matrix of $T$ is

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3
\end{pmatrix}_B.
$$
To get the standard matrix of $\iota^{-1}_B \circ T \circ \iota_B$, we begin by applying the transformation to the first standard basis vector. The basis vector $(1, 0, 0)$ is taken to $(1, 0, 0)_B$ by $\iota_B$, which is then taken to itself by $T$, then taken back to $(1, 0, 0)$ by $\iota_B^{-1}$. It follows that $\iota^{-1}_B \circ T \circ \iota_B$ takes $(1, 0, 0)$ to itself, and we have the first column of the standard matrix. The second standard basis vector $(0, 1, 0)$ is taken to $(0, 1, 0)_B$ by $\iota_B$, which is then taken to $(0, 2, 0)$ by $T$, then taken back to $(0, 2, 0)$ by $\iota_B^{-1}$, so

$$(\iota^{-1}_B \circ T \circ \iota_B)(0, 1, 0) = (0, 2, 0),$$

and we have the second column of the standard matrix. Finally,

$$(\iota^{-1}_B \circ T \circ \iota_B)(0, 0, 1) = \iota_B^{-1}(T((0, 0, 1)_B)) = \iota_B^{-1}((0, 0, 3)_B) = (0, 0, 3),$$

so the standard matrix of $\iota^{-1}_B \circ T \circ \iota_B$ is

$$
\begin{pmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3
\end{pmatrix}.
$$

**Problem 17** Assume $T : M_2 \to M_2$ is the transformation defined by

$$
T\left(\begin{pmatrix} a & b \\ c & d \end{pmatrix}\right) = \begin{pmatrix} d & c \\ b & a \end{pmatrix},
$$

and let

$$
\mathcal{B} = \{\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix}\}.
$$

Verify that the $\mathcal{B}$-matrix of $T$ has the same entries as the standard matrix of $\iota^{-1}_B \circ T \circ \iota_B$ by computing them both.

Commutative diagrams let us easily see a relationship that emerges from similar transformations. If $T_1$ and $T_2$ are similar transformations of a space $\mathcal{V}$, then by our definition of similar, there are two bases $\mathcal{B}$ and $\mathcal{C}$ so that the matrix entries in the $\mathcal{B}$-matrix of $T_1$ are the same as the matrix entries in the $\mathcal{C}$-matrix of $T_2$. In terms of the isomorphisms, this mouthful can be incorporated into a single diagram, which looks like

$$
\begin{array}{ccc}
\mathcal{V} & \xrightarrow{T_1} & \mathcal{V} \\
\iota_B & \uparrow & \iota_B^{-1} \\
R^n & \xrightarrow{\cong} & R^n \\
\iota_C & \downarrow & \iota_C^{-1} \\
\mathcal{V} & \xrightarrow{T_2} & \mathcal{V},
\end{array}
$$

which is a commutative diagram because the standard matrix of $\iota_B^{-1} \circ T_1 \circ \iota_B$ has the same entries as the $\mathcal{B}$-matrix of $T_1$, which has the same entries as the $\mathcal{C}$-matrix of $T_2$, which in turn has the same entries as the standard matrix of $\iota_C^{-1} \circ T_2 \circ \iota_C$. If we ignore the middle line, we are left with the diagram

$$
\begin{array}{ccc}
\mathcal{V} & \xrightarrow{T_1} & \mathcal{V} \\
\iota_C \circ \iota_B^{-1} & \downarrow & (\iota_C \circ \iota_B^{-1})^{-1} \\
\mathcal{V} & \xrightarrow{T_2} & \mathcal{V},
\end{array}
$$

from which we see that $T_1 = S^{-1}T_2S$, when $S = \iota_C \circ \iota_B^{-1}$. The converse of this observation is also true, and it can also be seen by drawing the appropriate commutative diagram: if $T_1 = S^{-1}T_2S$ for some invertible transformation $S$, then $T_1$ is similar to $T_2$.

**Example 11**

We have seen that the orthogonal projection of $R^2$ onto the $y$-axis is similar to the projection of $R^2$ onto the line $y = x$ in the direction of the line $y = 5x$. Find the standard matrix of an invertible transformation $S$ so that

$$
P_1 = S^{-1}P_2S,
$$

when $P_1$ and $P_2$ are these two projections. Verify the answer by multiplying out the standard matrices of these three transformations.
Assume that \( P_1 \) denotes the orthogonal projection of \( \mathbb{R}^2 \) onto the \( y \)-axis and \( P_2 \) denotes the projection of \( \mathbb{R}^2 \) onto the line \( y = x \) in the direction of the line \( y = 5x \). We need to find bases \( \mathcal{B} \) and \( \mathcal{C} \) so that the matrix entries of the \( \mathcal{B} \)-matrix of \( P_1 \) are the same as the matrix entries in the \( \mathcal{C} \)-matrix of \( T_2 \). We can force both matrices to look like
\[
\begin{pmatrix}
1 & 0 \\
0 & 0 \\
\end{pmatrix}
\]
if we choose bases of eigenvectors for these transformations, so let
\[
\mathcal{B} = \{(0, 1), (1, 0)\} \quad \text{and} \quad \mathcal{C} = \{(1, 1), (1, 5)\}.
\]
With this choice, the diagrams above will commute and the transformation \( S = \iota_{\mathcal{C}} \circ \iota_{\mathcal{B}}^{-1} \) will satisfy \( P_1 = S^{-1}P_2S \).

To get the standard matrix of \( S \), we need to compute \( S(1, 0) \) and \( S(0, 1) \), then write the standard coordinates of these outputs in the columns. To see what \( S \) does to \( (1, 0) \), we first need to know what the transformation \( \iota_{\mathcal{B}}^{-1} \) does to \( (1, 0) \), i.e. we need to find the vector \( (\alpha, \beta) \) so that \( \iota_{\mathcal{B}}(\alpha, \beta) = (1, 0) \). Since \( \iota_{\mathcal{B}}(\alpha, \beta) = (\alpha, \beta)_B \), this is exactly the same as finding the \( \mathcal{B} \)-coordinates of \( (1, 0) \);
\[
(1, 0) = \alpha(0, 1) + \beta(1, 0)
\]
implies \( \alpha = 0 \) and \( \beta = 1 \), and \( \iota_{\mathcal{B}}^{-1}(1, 0) = (0, 1) \). Now we have
\[
S(1, 0) = (\iota_{\mathcal{C}} \circ \iota_{\mathcal{B}}^{-1})(1, 0) = \iota_{\mathcal{C}}(0, 1) = (0, 1)_C = (0)(1, 1) + (1)(1, 5) = (1, 5).
\]
To get the second column of the standard matrix, we perform the similar computation on the second basis vector, getting
\[
S(0, 1) = (\iota_{\mathcal{C}} \circ \iota_{\mathcal{B}}^{-1})(0, 1) = \iota_{\mathcal{C}}(1, 0) = (1, 0)_C = (1)(1, 1) + (0)(1, 5) = (1, 1),
\]
and the standard matrix of \( S \) is
\[
\begin{pmatrix}
1 & 1 \\
5 & 1 \\
\end{pmatrix}.
\]
The standard matrices of \( P_1 \) and \( P_2 \) are
\[
\begin{pmatrix}
0 & 0 \\
0 & 1 \\
\end{pmatrix} \quad \text{and} \quad \begin{pmatrix}
\frac{5}{4} & -\frac{1}{4} \\
\frac{5}{4} & -\frac{1}{4} \\
\end{pmatrix},
\]
respectively. We can verify that \( P_1 = S^{-1}P_2S \) by checking the equivalent equation \( SP_1 = P_2S \), the left side of which is
\[
\begin{pmatrix}
1 & 1 \\
5 & 1 \\
\end{pmatrix} \begin{pmatrix}
0 & 0 \\
0 & 1 \\
\end{pmatrix} = \begin{pmatrix}
0 & 1 \\
0 & 1 \\
\end{pmatrix},
\]
and the right side is
\[
\begin{pmatrix}
\frac{5}{4} & -\frac{1}{4} \\
\frac{5}{4} & -\frac{1}{4} \\
\end{pmatrix} \begin{pmatrix}
1 & 1 \\
5 & 1 \\
\end{pmatrix} = \begin{pmatrix}
0 & 1 \\
0 & 1 \\
\end{pmatrix},
\]
which are equal.

**Problem 18** The orthogonal reflection of \( \mathbb{R}^2 \) across the \( x \)-axis is similar to the reflection of \( \mathbb{R}^2 \) across the line \( y = -x \) in the direction of the line \( y = 2x \). Find the standard matrix of an invertible transformation \( S \) so that
\[
P_1 = S^{-1}P_2S,
\]
when \( P_1 \) and \( P_2 \) are these two reflections. Verify the answer by multiplying out the standard matrices of these three transformations.
Example 12

Consider the transformations $P$ and $Q$ given by their standard matrices

$$
\begin{pmatrix}
1 & -1 \\
1 & -1
\end{pmatrix}
\quad \text{and} \quad
\begin{pmatrix}
0 & 1 \\
0 & 0
\end{pmatrix}
$$

(respectively). Let $S$ be the invertible transformation whose standard matrix is

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

Use the fact that $SP = QS$ to find two bases relative to which the matrices of $P$ and $Q$ have the same entries.

The fact that $SP = QS$ corresponds to saying that

$$
\begin{array}{c}
R^2 \xrightarrow{P} R^2 \\
S \downarrow \quad \downarrow S \\
R^2 \xrightarrow{Q} R^2
\end{array}
$$

is a commutative diagram. If $B = \{u, v\}$ denotes any basis of $R^2$, then

$$
\begin{array}{c}
R^2 \longrightarrow R^2 \\
\iota_B \downarrow \Downarrow \iota_B^{-1} \\
R^2 \xrightarrow{P} R^2
\end{array}
$$

is a commutative diagram, and we seek a second basis $C$ so that

$$
\begin{array}{c}
R^2 \xrightarrow{P} R^2 \\
\iota_B \Downarrow \iota_C \\
R^2 \xrightarrow{Q} R^2
\end{array}
$$

commutes. Our problem is then, how should we combine the first two diagrams to discover what $\iota_C$ should be? The diagrams combine into the commutative conglomerate

$$
\begin{array}{c}
R^2 \longrightarrow R^2 \\
\iota_B \Downarrow \iota_B^{-1} \\
S \downarrow \quad \downarrow S^{-1} \\
R^2 \xrightarrow{Q} R^2
\end{array}
$$

which strongly suggests that $\iota_C$ should be $S \circ \iota_B$. Now we figure out what $C$ needs to be to yield such an isomorphism $\iota_C = S \circ \iota_B$:

$$
\iota_C(\alpha, \beta) = S(\iota_B(\alpha, \beta)) = S(\alpha u + \beta v) = \alpha S(u) + \beta S(v)
$$

shows that we should take $C = \{S(u), S(v)\}$, and we will then have the entries of the $C$-matrix of $Q$ equal to the entries of the $B$-matrix of $P$.

Problem 19

In the solution of example 12, assume that $B = \{(1, 1), (0, 1)\}$ and $C = \{S(1, 1), S(0, 1)\}$. Compute the $B$-matrix of $P$ and the $C$-matrix of $Q$. 
Problem 20  Let $T$ denote the transformation whose standard matrix is

\[
\begin{pmatrix}
  0 & 1 \\
  0 & 0
\end{pmatrix},
\]

and let $S$ denote the invertible transformation with standard matrix

\[
\begin{pmatrix}
  1 & 2 \\
  0 & 3
\end{pmatrix}.
\]

Find two bases relative to which the matrices of $T$ and $STS^{-1}$ have the same entries.

More Exercises

1. polynomial interpolation and vandermond matrices
Chapter 8

Canonical Forms

A **canonical form** is a natural matrix form for a linear transformation. Usually, the canonical form is a complete similarity invariant: two transformations are similar exactly when the canonical matrix forms of the two are the same. There are two standard canonical forms that are widely used in linear algebra, the **Jordan canonical form** and the **rational canonical form**. The Jordan canonical form for a linear transformation may be found in several elementary textbooks on linear algebra, while the rational form is found in very few, primarily because the rational form requires advanced methods to rigorously establish. Both are extremely useful, but each form has advantages and disadvantages. When a question about a linear transformation arises, viewing the transformation from the perspective of one of these natural forms often makes the answer transparent.

### 8.1 Jordan form

Diagonalization is a nearly optimal matrix form for a linear transformation, but not every transformation can be diagonalized. There are two obstructions to diagonalization for a transformation, one of which involves a deficiency with the field of scalars, and the other is something intrinsic to the transformation itself. For example, a rotation of the plane by 90 degrees is not diagonalizable, and this is a consequence of the fact that its characteristic polynomial, which is $x^2 + 1$, has no real roots, but when we consider the transformation of $C^2$ with exactly the same standard matrix, we then find it has a diagonal form, corresponding to the fact that $x^2 + 1$ factors over $C$ as $(x - i)(x + i)$. This shows that the obstruction to the diagonalization of a rotation corresponds to a property of the field of real numbers, namely, that there are nonconstant real polynomials with no real roots. By way of contrast, consider the transformation of $R^2$ whose standard matrix is

\[
\begin{pmatrix}
0 & 1 \\
0 & 0
\end{pmatrix}.
\]

This transformation fails to be diagonalizable, but the nondiagonalizability has nothing to do with the scalar field. The transformation of $C^2$ with the same standard matrix is equally nondiagonalizable.

The **fundamental theorem of algebra** asserts that nonconstant complex polynomials always have a complex root, so by restricting our attention to complex vector spaces we eliminate one source of nondiagonalizability. The Jordan form is given to a transformation in this restricted setting, and the Jordan form of a transformation is its diagonal form when the transformation is diagonalizable. The fact is that every transformations of a complex space assumes a block diagonal form with blocks that look like

\[
\begin{pmatrix}
\beta & 1 \\
0 & \beta
\end{pmatrix}
\begin{pmatrix}
\gamma & 1 & 0 \\
0 & \gamma & 1 \\
0 & 0 & \gamma
\end{pmatrix} \cdots
\begin{pmatrix}
\omega & 1 & 0 & \cdots & 0 \\
0 & \omega & 1 & \cdots & 0 \\
\vdots & \ddots & \ddots & \ddots & \vdots \\
0 & \cdots & 0 & 1 \\
0 & \cdots & 0 & 0 & \omega
\end{pmatrix},
\]
where the scalar entries on the diagonals of the blocks constitute the eigenvalues of the transformation. These blocks are called **Jordan blocks**, and the block diagonal Jordan form gives an immediate idea of how close to being diagonalizable the transformation is. The diagonalizable transformations have only $1 \times 1$ Jordan blocks in their Jordan form, while the presence of one or more larger Jordan block moves the transformation away from diagonalizability.

If we are given a $2 \times 2$ matrix and we are asked for the Jordan form, the answer will always be one of two things; the matrix is either diagonalizable, in which the Jordan form looks like

$$
\begin{pmatrix}
a & 0 \\
0 & b
\end{pmatrix}
$$

with $a$ and $b$ the eigenvalues, or the matrix is not diagonalizable in which case it is similar to a matrix of the form

$$
\begin{pmatrix}
a & 1 \\
0 & a
\end{pmatrix},
$$

where $a$ is the lonely eigenvalue.

**Example 1**

Find the Jordan form of

$$
\begin{pmatrix}
1 & -4 \\
1 & 1
\end{pmatrix},
$$

When someone asks us to find the Jordan form of a matrix, they are asking for the Jordan form for the transformation whose standard matrix is the one given. The first step is to find the eigenvalues; computing the characteristic polynomial gives us

$$(1 - x)^2 + 4 = x^2 - 2x + 3,$$

and the quadratic formula reveals the roots, which are

$$
\frac{2 + \sqrt{4 - (4)(3)}}{2} \text{ and } \frac{2 - \sqrt{4 - (4)(3)}}{2},
$$

i.e. the roots are $1 + \sqrt{2}i$ and $1 - \sqrt{2}i$. Since the two distinct roots will yield two independent eigenvectors, the matrix is diagonalizable and similar to

$$
\begin{pmatrix}
1 + \sqrt{2}i & 0 \\
0 & 1 - \sqrt{2}i
\end{pmatrix},
$$

which is the Jordan form.

The Jordan form is not unique in the sense that the placement of the eigenvalues on the diagonal is not specified. If one is to check to see if two Jordan forms are the same, and both are diagonal matrices, one should check that the same eigenvalues are on the diagonal (but not necessarily in the same place). A perfectly legitimate answer to example 1 would have been to say that the Jordan form is

$$
\begin{pmatrix}
1 - \sqrt{2}i & 0 \\
0 & 1 + \sqrt{2}i
\end{pmatrix}.
$$

**Problem 1** Find the Jordan form of

$$
\begin{pmatrix}
2 & 1 \\
1 & 2
\end{pmatrix}.
$$

**Problem 2** Find the Jordan form of

$$
\begin{pmatrix}
0 & -1 \\
1 & 0
\end{pmatrix}.$$
Example 2

Find the Jordan form of
\[
\begin{pmatrix}
3 & -1 \\
1 & 1
\end{pmatrix}.
\]

We find that the characteristic polynomial is \(x^2 - 4x + 4\), so the only eigenvalue is 2. The Jordan form is then either
\[
\begin{pmatrix}
2 & 0 \\
0 & 2
\end{pmatrix} \quad \text{or} \quad \begin{pmatrix}
2 & 1 \\
0 & 2
\end{pmatrix},
\]
and since we were not given a scalar transformation, the Jordan form must be
\[
\begin{pmatrix}
2 & 1 \\
0 & 2
\end{pmatrix}.
\]

Here is a way to convince ourselves that the transformation in example 2 is similar to its Jordan form. If we select a basis using an eigenvector of
\[
\begin{pmatrix}
3 & -1 \\
1 & 1
\end{pmatrix}
\]
as the first basis vector, then the corresponding \(B\)-matrix will be of the form
\[
\begin{pmatrix}
2 & a \\
0 & b
\end{pmatrix}_B.
\]
Knowing that the only eigenvalue is 2 tells us that \(b = 2\), and since the transformation is not a scalar transformation we know that \(a \neq 0\). Thus the matrix is similar to
\[
\begin{pmatrix}
2 & a \\
0 & 2
\end{pmatrix},
\]
which is itself similar to
\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}
\begin{pmatrix}
2 & a \\
0 & 2
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix} = \begin{pmatrix}
2 & 1 \\
0 & 2
\end{pmatrix}.
\]

The methods used in this computation combine the two ways of thinking about similarity: first we changed basis, then we multiplied on both sides by an invertible matrix and its inverse. The computation really shows that any matrix like
\[
\begin{pmatrix}
2 & a \\
0 & 2
\end{pmatrix}
\]
with \(a \neq 0\) is similar to any other matrix of the same form with a different non-zero value for \(a\).

Problem 3

Find the Jordan form of
\[
\begin{pmatrix}
1 & 0 \\
7 & 1
\end{pmatrix}.
\]

Problem 4

Find the Jordan form of
\[
\begin{pmatrix}
1 & 0 \\
0 & 1
\end{pmatrix}.
\]
Problem 5  Assume $A$ denotes the matrix given in problem 3 and $J$ denotes its Jordan form. Find an invertible matrix $S$ so that $SA = JS$.

When we move to three dimensions the situation becomes slightly more complicated. A $3 \times 3$ complex matrix is always similar to one of the following types:

\[
\begin{pmatrix}
  a & 0 & 0 \\
  0 & b & 0 \\
  0 & 0 & c
\end{pmatrix}
\quad \text{or} \quad
\begin{pmatrix}
  a & 1 & 0 \\
  0 & a & 0 \\
  0 & 0 & b
\end{pmatrix}
\quad \text{or} \quad
\begin{pmatrix}
  a & 1 & 0 \\
  0 & a & 1 \\
  0 & 0 & a
\end{pmatrix}.
\]

Notice that the diagonalizable ones are given by the leftmost matrix, and the non-diagonalizable ones are now of two distinct types, those with only one dimension of eigenvectors (the rightmost matrix), and those with two dimensions of eigenspaces (the center matrix).

Problem 6  Find the Jordan forms of the following matrices:

i. \[
\begin{pmatrix}
  3 & 0 & 0 \\
  0 & 3 & 0 \\
  0 & 0 & 3
\end{pmatrix}
\]

ii. \[
\begin{pmatrix}
  0 & -1 & 0 \\
  1 & 0 & 0 \\
  0 & 0 & i
\end{pmatrix}
\]

iii. \[
\begin{pmatrix}
  0 & 1 & 3 \\
  0 & 0 & 2 \\
  0 & 0 & 0
\end{pmatrix}
\]

iv. \[
\begin{pmatrix}
  4 & -1 & 0 \\
  1 & 2 & 0 \\
  0 & 0 & 1
\end{pmatrix}
\]

Problem 7  Make an educated guess as to what possible types of Jordan forms are obtainable in four dimensions.

8.2 Rational form

In order to freely express transformations in a Jordan form, the field of scalars must behave like the complex numbers; nonconstant polynomials must be expressible as a product of linear factors. The rational form for a transformation has no such limitation. No matter what the field of scalars is, every transformation will have a corresponding matrix in rational form, and the entries of this matrix lie within the scalar field. Being able to attain a canonical form without leaving the field of scalars is the principal advantage of the rational form over the Jordan form.

Assume $T$ is a linear transformation that maps a vector space into itself, and suppose $u$ is a nonzero vector in that space. The set $\{u\}$ is independent, and if $u$ is not an eigenvector of $T$, then the set $\{u, T(u)\}$ is also independent. It might happen that $\{u, T(u), T(T(u))\}$ is even independent, and some transformations have the property that there exists a vector $u$ for which it is possible to build a basis of the vector space using this idea; if $u$ is a vector in the domain of $T$ for which

\[
\{u, T(u), \ldots, T^{n-1}(u)\}
\]

is a basis of the domain, then we call $T$ a cyclic transformation and we refer to $u$ as a cyclic vector for $T$. The matrix of $T$ relative to this basis is particularly simple, and the form of the resulting matrix provides the basic building block for the rational canonical form.

Problem 8  Assume $T$ is the transformation of $R^3$ whose standard matrix is

\[
\begin{pmatrix}
  1 & 0 & 0 \\
  0 & 2 & 0 \\
  0 & 0 & 3
\end{pmatrix}.
\]

Let $u = (1, 1, 1)$ and $B = \{u, T(u), T^2(u)\}$. Find the $B$-matrix of $T$.

Problem 9  Assume $T$ is an arbitrary transformation of $R^3$ and $B = \{u, T(u), T^2(u)\}$ is a basis. Find the first two columns of the $B$-matrix of $T$. 
For transformations of the plane, virtually every one is cyclic. In order for a transformation $T$ of $R^2$ to fail to be cyclic, it must be that $\{u, T(u)\}$ is never independent, which is equivalent to saying that every nonzero vector is an eigenvector of $T$. The only transformations with this property are the scalar transformations. When $T$ is a nonscalar transformation and $u$ is not an eigenvector of $T$, we will have that $B = \{u, T(u)\}$ is a basis of $R^2$, and the corresponding $B$-matrix of $T$ is called the rational canonical form of $T$.

**Example 3**

Assume $T : R^2 \to R^2$ has standard matrix

$$
\begin{pmatrix}
2 & 3 \\
0 & 5
\end{pmatrix}.
$$

Find the rational form for $T$.

We begin by searching for a vector that is not an eigenvector for $T$. The first column of the standard matrix for $T$ shows that we should avoid the first basis vector, since it is an eigenvector, and the second column of the matrix tells us that the second basis vector is not an eigenvector. Thus we let $u = (0, 1)$, and we get $B = \{u, T(u)\} = \{(0,1), (3,5)\}$, relative to which, the $B$-matrix of $T$ is

$$
\begin{pmatrix}
0 & -10 \\
1 & 7
\end{pmatrix}_B,
$$

which is the rational form for $T$.

There are infinitely many cyclic vector for the transformation $T$ in example 3. Besides the vector $(0, 1)$, we could have taken $u$ to be $(1,2)$ or $(-1,1)$, because neither of these is an eigenvector for $T$. If we were to let $B = \{u, T(u)\} = \{(1,2), (8,10)\}$ or $B = \{u, T(u)\} = \{(-1,1), (1,5)\}$, we would still get the same $B$-matrix as we got above. Why should this be?

**Problem 10** Assume $T : R^2 \to R^2$ has standard matrix

$$
\begin{pmatrix}
2 & 0 \\
0 & 5
\end{pmatrix}.
$$

Find the rational form for $T$.

**Problem 11** Find the determinants of the following matrices. Compare the answers to the entries in the matrices.

\[(i) \begin{pmatrix} 0 & 1 \\ 1 & 2 \end{pmatrix} \quad (ii) \begin{pmatrix} 0 & 5 \\ 1 & -1 \end{pmatrix} \quad (iii) \begin{pmatrix} 0 & -10 \\ 1 & -6 \end{pmatrix} \quad (iv) \begin{pmatrix} 0 & c \\ 1 & b \end{pmatrix}\]

Because of the relationship between the entries of the matrix

$$
\begin{pmatrix}
1 & a_0 \\
& a_1 \\
& \ddots \\
& & \ddots \\
& & & 1 & a_{n-1}
\end{pmatrix}
$$

(all the missing entries are zero) and the coefficients of its characteristic polynomial, which is

$$x^n - a_{n-1}x^{n-1} - \ldots - a_1x - a_0,$$

people refer to a matrix of this form as the companion matrix of the polynomial. Thus, given any polynomial $p$, writing down the companion matrix of $p$ lets us construct a cyclic transformation whose characteristic polynomial is $p$. 
Problem 12 Give an example of a matrix whose characteristic polynomial is \( x^3 + 2x^2 - 5x + 4 \).

The Jordan Form of a transformation is a block diagonal matrix where each block is a Jordan block, i.e. a block of the form

\[
\begin{pmatrix}
a & 1 \\
a & 1 \\
\vdots & \vdots \\
a & 1 \\
a & a
\end{pmatrix}.
\]

Thus examples of \( 1 \times 1, 2 \times 2, 3 \times 3 \) and \( 4 \times 4 \) Jordan Blocks are

\[
(13), \quad \begin{pmatrix} 3 & 1 \\ 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 23 & 1 & 0 \\ 0 & 23 & 1 \\ 0 & 0 & 23 \end{pmatrix}, \quad \begin{pmatrix} \pi & 1 & 0 & 0 \\ 0 & \pi & 1 & 0 \\ 0 & 0 & \pi & 1 \\ 0 & 0 & 0 & \pi \end{pmatrix},
\]

and an example of a typical Jordan form might be

\[
\begin{pmatrix}
J_1 & 0 & 0 & 0 \\
0 & J_2 & 0 & 0 \\
0 & 0 & J_3 & 0 \\
0 & 0 & 0 & J_4
\end{pmatrix}
\]

where the four blocks \( J_1, J_2, J_3, \) and \( J_4 \) are the four Jordan blocks above.

The Rational Form of a transformation is a block diagonal matrix where each block is a rational block, i.e. a block of the form

\[
\begin{pmatrix}
0 & \ldots & a_0 \\
1 & \ldots & a_1 \\
\vdots & \ddots & \vdots \\
0 & \ldots & a_{n-2} \\
1 & \ldots & a_{n-1}
\end{pmatrix}.
\]

Examples of \( 1 \times 1, 2 \times 2, 3 \times 3 \) and \( 4 \times 4 \) rational blocks are

\[
(13), \quad \begin{pmatrix} 0 & 2 \\ 1 & 3 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 4 \\ 1 & 0 & 1 \\ 0 & 1 & 6 \end{pmatrix}, \quad \begin{pmatrix} 0 & 0 & 0 & 8 \\ 1 & 0 & 0 & 4 \\ 0 & 1 & 0 & 2 \\ 0 & 0 & 1 & 5 \end{pmatrix}.
\]

A rational block is the matrix for a transformation \( T \) relative to a basis of the form

\[\{u, T(u), T^2(u), \ldots, T^{n-1}(u)\}\].

A transformation is cyclic exactly when its rational form consists of a single rational block. The rational form of a general transformation looks like

\[
\begin{pmatrix}
R_1 \\
\vdots \\
R_k
\end{pmatrix},
\]

where each \( R_1, \ldots, R_k \) is a rational block. The more blocks that appear in a rational form, the further away from being cyclic is the corresponding transformation.

If we are given a bunch of Jordan blocks, we could use them to construct a block diagonal matrix that will be the Jordan form of the corresponding transformation. If someone else were to use the same blocks to construct a block diagonal matrix, putting the blocks in possibly different positions
on the diagonal, we would consider the resulting block matrix the same Jordan form. Thus, the Jordan form is determined by which blocks are used, not the position on the diagonal where they appear. By contrast, if we are given a bunch of rational blocks, then the block diagonal matrix obtained by placing these blocks along a diagonal is quite possibly not the rational form of any transformation. The rational form imposes a condition that relates the blocks to one another; if a rational block with characteristic polynomial \( p \) is placed on the diagonal, any new block added below this one must have its characteristic polynomial divide \( p \). Thus, the matrices

\[
\begin{pmatrix}
3 & 0 \\
0 & 3
\end{pmatrix}
\begin{pmatrix}
0 & -2 & 0 \\
1 & 3 & 0 \\
0 & 0 & 2
\end{pmatrix}
\begin{pmatrix}
0 & -2 & 0 \\
1 & 3 & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
0 & -2 & 0 & 0 \\
1 & 3 & 0 & 0 \\
0 & 0 & 0 & -2 \\
0 & 0 & 1 & 3
\end{pmatrix}
\]

are legitimate rational forms, while the rational block diagonal matrices

\[
\begin{pmatrix}
3 & 0 \\
0 & 2
\end{pmatrix}
\begin{pmatrix}
0 & -2 & 0 \\
1 & 3 & 0 \\
0 & 0 & 3
\end{pmatrix}
\begin{pmatrix}
2 & 0 & 0 \\
0 & 0 & -2 \\
0 & 1 & 3
\end{pmatrix}
\begin{pmatrix}
0 & -2 & 0 & 0 \\
1 & 3 & 0 & 0 \\
0 & 0 & 0 & -3 \\
0 & 0 & 1 & 4
\end{pmatrix}
\]

are not.

**Example 4**

Consider the transformation whose standard matrix is

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 2 & 0 \\
0 & 0 & 3
\end{pmatrix}
\]

Give three other rational block diagonal matrices that are similar to this one, and indicate what the rational form is.

If we replace the \( 2 \times 2 \) northwest block with its rational form, we obtain the matrix

\[
\begin{pmatrix}
0 & -2 & 0 \\
1 & 3 & 0 \\
0 & 0 & 3
\end{pmatrix}
\]

which is similar to the matrix we started with. We could also replace the \( 2 \times 2 \) southeast block with its rational form, and obtain either

\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 0 & -6 \\
0 & 1 & 5
\end{pmatrix}
\]

\[
\begin{pmatrix}
0 & -6 & 0 \\
1 & 5 & 0 \\
0 & 0 & 1
\end{pmatrix}
\]

depending on where the blocks are placed on the diagonal. We now have four matrices similar to one another, but none of these is the rational form because none of them satisfies the condition that binds the diagonal blocks together. The matrix we started with is actually cyclic, with cyclic vector \((1, 1, 1)\), and the matrix of this transformation relative to the basis \(\{(1, 1, 1), (1, 2, 3), (1, 4, 9)\}\) is

\[
\begin{pmatrix}
0 & 0 & 6 \\
1 & 0 & -11 \\
0 & 1 & 6
\end{pmatrix}
\]

which is the rational form.
Problem 13 Consider the transformation whose standard matrix is
\[
\begin{pmatrix}
-1 & 0 & 0 \\
0 & 5 & 0 \\
0 & 0 & 2
\end{pmatrix}.
\]
Give three other rational block diagonal matrices that are similar to this one, and indicate what the rational form is.

It is possible to derive the rational form from the Jordan form. The pertinent theorem says that a transformation is cyclic if and only if its Jordan form contains exactly one Jordan block corresponding to each distinct eigenvalue. Thus the matrices
\[
\begin{pmatrix}
1 & 0 \\
0 & 3
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 \\
0 & 2 & 0 \\
0 & 0 & 1
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 8 & 1 \\
0 & 0 & 0 & 8
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 & 0 & 0 \\
0 & 2 & 0 & 0 & 0 \\
0 & 0 & 8 & 0 & 0 \\
0 & 0 & 0 & 9 & 0 \\
0 & 0 & 0 & 0 & 3
\end{pmatrix}
\]
correspond to cyclic transformations, while
\[
\begin{pmatrix}
3 & 0 \\
0 & 3
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 \\
0 & 2 & 0 \\
0 & 0 & 2
\end{pmatrix}
\begin{pmatrix}
2 & 1 & 0 & 0 \\
0 & 2 & 0 & 0 \\
0 & 0 & 2 & 1 \\
0 & 0 & 0 & 2
\end{pmatrix}
\begin{pmatrix}
8 & 1 & 0 & 0 & 0 \\
0 & 8 & 0 & 0 & 0 \\
0 & 0 & 7 & 0 & 0 \\
0 & 0 & 0 & 9 & 0 \\
0 & 0 & 0 & 0 & 8
\end{pmatrix}
\]
do not. The way to get the rational form from the Jordan form is to select exactly one of the largest Jordan blocks for each distinct eigenvalue and shuffle the diagonal blocks so that the selected ones occupy the northwest corner. Taken together, these selected blocks correspond to a single rational block. One then repeats the process with the remaining Jordan blocks until they have all been exhausted.

Example 5 Find the rational form for the transformation whose Jordan form has the following Jordan blocks along its diagonal.
\[
\begin{pmatrix}
9 & 8 \\
0 & 8
\end{pmatrix}
\begin{pmatrix}
8 & 1 \\
0 & 8
\end{pmatrix}
\begin{pmatrix}
8 & 1 \\
0 & 8
\end{pmatrix}
\begin{pmatrix}
9 & 1 \\
0 & 9
\end{pmatrix}
\]
The distinct eigenvalues are 7, 8, and 9, so we select exactly one of the largest blocks corresponding to each distinct eigenvalue, gathering the matrices
\[
\begin{pmatrix}
7 & 8 & 9 \\
0 & 1 & 1
\end{pmatrix}
\]
and leaving
\[
\begin{pmatrix}
9 & 8 \\
0 & 8
\end{pmatrix}
\begin{pmatrix}
8 & 1 \\
0 & 8
\end{pmatrix}
\begin{pmatrix}
9 & 1 \\
0 & 9
\end{pmatrix}
\]
The 5 × 5 matrix that results from the 3 selected Jordan blocks is cyclic, with characteristic polynomial
\[(x - 7)(x - 8)^2(x - 9)^2,
\]
which may be rewritten as
\[x^5 - 41x^4 + 671x^3 - 5479x^2 + 22320x - 36288,
\]
and the companion matrix of this polynomial is
\[
\begin{pmatrix}
0 & 0 & 0 & 36288 \\
1 & 0 & 0 & -22320 \\
0 & 1 & 0 & 5479 \\
0 & 0 & 1 & 41
\end{pmatrix}, \quad \text{which is the rational form of}
\begin{pmatrix}
7 & 0 & 0 & 0 \\
0 & 8 & 1 & 0 \\
0 & 0 & 8 & 0 \\
0 & 0 & 0 & 9
\end{pmatrix}.
\]
We now repeat the process with the three blocks left over, beginning by selecting one of the largest blocks corresponding to each distinct eigenvalue. The blocks left over were
\[(9) \quad (8) \quad \begin{pmatrix} 8 & 1 \\ 0 & 8 \end{pmatrix},\]
so there are now the two distinct eigenvalues 8 and 9, and we select the blocks
\[(9) \quad \text{and} \quad \begin{pmatrix} 8 & 1 \\ 0 & 8 \end{pmatrix}.
\]
The characteristic polynomial of
\[
\begin{pmatrix}
9 & 0 & 0 \\
0 & 8 & 1 \\
0 & 0 & 8
\end{pmatrix}
\]
is
\[(x - 9)(x - 8)^2 = x^3 - 25x^2 + 208x - 576,
\]
and the rational form for this cyclic block is the corresponding companion matrix
\[
\begin{pmatrix}
0 & 0 & 576 \\
1 & 0 & -208 \\
0 & 1 & 25
\end{pmatrix}.
\]
The last block left is the $1 \times 1$ matrix $(8)$, which is already in both its Jordan and its rational form. We now obtain the rational form for the transformation whose Jordan blocks were given by writing our rational block on the diagonal, so that the characteristic polynomials of the lower blocks divide those of the ones above them, like
\[
\begin{pmatrix}
0 & 0 & 0 & 0 & 36288 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & -22320 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 5479 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & -671 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 41 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 576 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & -208 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 25 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 8
\end{pmatrix}.
\]

**Problem 14** Find the rational form for the transformation whose Jordan form has the following Jordan blocks along its diagonal.
\[
\begin{pmatrix}
9 \\
0
\end{pmatrix} \quad \begin{pmatrix} 8 & 1 \\ 0 & 8 \end{pmatrix} \quad \begin{pmatrix} 7 & 1 \\ 0 & 7 \\ 0 & 0 \end{pmatrix} \quad \begin{pmatrix} 8 & 1 \\ 0 & 8 \end{pmatrix} \quad \begin{pmatrix} 7 & 1 \\ 0 & 7 \end{pmatrix}
\]

**Problem 15** Find the rational forms of the transformations whose standard matrices are the following.
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 3
\end{pmatrix} \quad \begin{pmatrix} 1 & 1 \\ 1 & 1 \\ 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \begin{pmatrix} 1 & -1 & 0 & 0 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix} \quad \begin{pmatrix} 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix}
\]

More Exercises
Chapter 9

Unitary Geometry

9.1 Standard inner product

The notion of similarity lets us obtain canonical forms for arbitrary linear transformations on finite dimensional spaces by carefully selecting a suitable basis. The generality is the fundamental strength of the concept, but the generality comes with a cost; we are often led to a basis that fails to have the familiar properties we are used to with the standard basis. The standard basis consists of vectors of length one, each vector meeting at a right angle to all other vectors in the basis. We call this an orthonormal basis, and one can imagine other orthonormal bases by rotating the standard basis. A variant of the similarity concept is the idea of unitary equivalence; two transformations are unitarily equivalent when they share a common matrix form relative to two orthonormal bases. The advantage we gain is the ability to work with manageable bases, but the cost is a loss of generality. There is no known powerful canonical form for unitary equivalence, like the Jordan and rational forms give us for similarity. On the other hand, it turns out to be surprisingly easy to recognize the transformations that can be diagonalized with an orthonormal basis. The theorem that tells us which transformations can be orthogonally diagonalized is called the spectral theorem, and this is the principle result about unitary equivalence.

The algebraic gadget that carries geometric information, like orthogonality and length, is the inner product (also called a scalar product or dot product). The notation used to denote an inner product applied to a pair of vectors \( \mathbf{u} \) and \( \mathbf{v} \) looks like

\[
\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{u} | \mathbf{v} \rangle = [\mathbf{u}, \mathbf{v}] = \langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u} \cdot \mathbf{v} = \mathbf{v}^* \mathbf{u},
\]

depending on which book one reads. In this book we will use the first and last notation.

If \( \mathbf{u} \) and \( \mathbf{v} \) are elements of \( \mathbb{C}^n \) with standard coordinates

\[
\mathbf{u} = (\alpha_1, \ldots, \alpha_n) \quad \text{and} \quad \mathbf{v} = (\beta_1, \ldots, \beta_n),
\]

then the standard inner product is defined by

\[
\langle \mathbf{u}, \mathbf{v} \rangle = \alpha_1 \beta_1 + \ldots + \alpha_n \beta_n,
\]

where \( \beta \) denotes the complex conjugate of the complex number \( \beta \). When \( \mathbf{u} \) and \( \mathbf{v} \) are both in \( \mathbb{R}^n \), this definition reduces to

\[
\langle \mathbf{u}, \mathbf{v} \rangle = \alpha_1 \beta_1 + \ldots + \alpha_n \beta_n,
\]

since \( \bar{\beta} = \beta \) when \( \beta \) is a real number.

Until now we have made no distinction between a vector whose coordinates are listed horizontally or vertically, but it is convenient to do so now so that the inner product can be viewed as a matrix
product. We will henceforth write the coordinates of vectors vertically, so that refering to the
standard coordinates of \( \mathbf{u} \), as above, now means

\[
\mathbf{u} = \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix},
\]

and we define the **transpose of a vector** \( \mathbf{u} \) to be the horizontal list of coordinates

\[
\mathbf{u}^t = (\alpha_1, \ldots, \alpha_n).
\]

We also define the **conjugate transpose** or **adjoint** of \( \mathbf{u} \) to be

\[
\mathbf{u}^* = (\bar{\alpha}_1, \ldots, \bar{\alpha}_n).
\]

With these definitions, the inner product of \( \mathbf{u} \) and \( \mathbf{v} \) takes the form

\[
\langle \mathbf{u}, \mathbf{v} \rangle = (\bar{\beta}_1, \ldots, \bar{\beta}_n) \begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix} = v^* u.
\]

To conserve page space, we may occasionally refer to the coordinates a vector \( \mathbf{u} \) by giving the
horizontal list for \( \mathbf{u}^t \).

### 9.2 Fourier expansion

Viewing the inner product as a matrix multiplication shows us that the inner product is linear in
its first variable and **conjugate linear** in its second variable, i.e. for every \( \mathbf{u} \) and for every linear
combination of the \( \mathbf{v} \)'s we have

\[
\langle \mathbf{u}, \alpha_1 \mathbf{v}_1 + \alpha_2 \mathbf{v}_2 \rangle = \bar{\alpha}_1 \langle \mathbf{u}, \mathbf{v}_1 \rangle + \bar{\alpha}_2 \langle \mathbf{u}, \mathbf{v}_2 \rangle.
\]

The inner product of a vector with itself returns the length of the vector squared, since

\[
\langle \mathbf{u}, \mathbf{u} \rangle = \mathbf{u}^* \mathbf{u} = \alpha_1 \bar{\alpha}_1 + \ldots + \alpha_n \bar{\alpha}_n.
\]

When two arrows with joined tails form a right angle, the inner product of the corresponding vectors
is zero, and the converse is equally true, so the inner product detects orthogonality by the condition
\( \langle \mathbf{u}, \mathbf{v} \rangle = 0 \) if and only if \( \mathbf{u} \) is orthogonal to \( \mathbf{v} \). When \( \mathcal{B} \) is an orthonormal basis, these three
properties of the inner product let us see what the \( \mathcal{B} \)-coordinates of a vector are without having to
compute them directly; if \( \mathcal{B} = \{\mathbf{e}_1, \ldots, \mathbf{e}_n\} \) and

\[
\mathbf{u} = \alpha_1 \mathbf{e}_1 + \ldots + \alpha_n \mathbf{e}_n,
\]

the fact that the vectors in \( \mathcal{B} \) have length one says \( \langle \mathbf{e}_i, \mathbf{e}_i \rangle = 1 \), and the orthogonality of these
vectors says that \( \langle \mathbf{e}_i, \mathbf{e}_j \rangle = 0 \) whenever \( i \neq j \), so by taking the inner product of both sides of
the equation with \( \mathbf{e}_i \), and using the linearity, we prove that \( \alpha_i = \langle \mathbf{u}, \mathbf{e}_i \rangle \). Thus we may view the
\( \mathcal{B} \)-coordinates of \( \mathbf{u} \) by writing

\[
\mathbf{u} = \langle \mathbf{u}, \mathbf{e}_1 \rangle \mathbf{e}_1 + \ldots + \langle \mathbf{u}, \mathbf{e}_n \rangle \mathbf{e}_n,
\]

and this is called the **Fourier expansion** of \( \mathbf{u} \) relative to the orthonormal basis \( \mathcal{B} \).
Example 1

Assume \( B = \{e_1, e_2, e_3\} \) with \( e_1 = (0, 1, 0), e_2 = (\sqrt{3}/2, 0, \sqrt{3}/2), \) and \( e_3 = (\sqrt{3}/2, 0, -\sqrt{3}/2). \) Given the Fourier expansion of the vector \( u \) relative to \( B \) when \( u' = (1, 1, 1), \) and if the standard matrix of \( T \) is

\[
\begin{pmatrix}
0 & 1 & 3 \\
1 & 1 & 1 \\
3 & 1 & 0
\end{pmatrix},
\]

find the \( B \)-matrix of \( T. \)

We begin by finding the \( B \)-coordinates of \( u, \) which are \( \langle u, e_1 \rangle, \langle u, e_2 \rangle \) and \( \langle u, e_3 \rangle. \) Computing these inner products, we find

\[
\langle u, e_1 \rangle = e_1^* u = (1)(0) + (1)(1) + (1)(0) = 1,
\]

\[
\langle u, e_2 \rangle = e_2^* u = (1)(\sqrt{3}/2) + (1)(0) + (1)(\sqrt{3}/2) = \sqrt{2},
\]

and

\[
\langle u, e_3 \rangle = e_3^* u = (1)(\sqrt{3}/2) + (1)(0) + (1)(-\sqrt{3}/2) = 0.
\]

It follows that

\[
u = e_1 + \sqrt{2}e_2,
\]

which is the Fourier expansion of \( u \) relative to \( B. \)

To get the first column of the \( B \)-matrix of \( T, \) we need to compute the \( B \)-coordinates of \( T(e_1). \) Since \( e_1 \) is the second standard basis vector, we see the standard coordinates of \( T(e_1) \) in the second column of the standard matrix for \( T, \) where we see \( T(e_1) = u \) with \( u \) as above. There we found that the \( B \)-coordinates are \( 1, \sqrt{2}, \) and \( 0, \) so the \( B \)-matrix of \( T \) begins to take shape as

\[
\begin{pmatrix}
1 & ? & ? \\
\sqrt{2} & ? & ? \\
0 & ? & ?
\end{pmatrix}.
\]

The second column of the \( B \)-matrix contains the \( B \)-coordinates of the vector \( T(e_2), \) which are the scalars \( \langle T(e_2), e_1 \rangle, \langle T(e_2), e_2 \rangle \) and \( \langle T(e_2), e_3 \rangle. \) Now

\[
T(e_2) = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 1 & 1 \\ 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{3}/2 \\ 0 \\ \sqrt{3}/2 \end{pmatrix} = \begin{pmatrix} 3\sqrt{3}/2 \\ \sqrt{2} \\ 3\sqrt{3}/2 \end{pmatrix},
\]

so

\[
\langle T(e_2), e_1 \rangle = e_1^* T(e_2) = \begin{pmatrix} 3\sqrt{3}/2 \\ \sqrt{2} \\ 3\sqrt{3}/2 \end{pmatrix} = \sqrt{2},
\]

\[
\langle T(e_2), e_2 \rangle = e_2^* T(e_2) = \begin{pmatrix} \sqrt{3}/2 \\ 0 \\ 3\sqrt{3}/2 \end{pmatrix} = 3,
\]

and

\[
\langle T(e_2), e_3 \rangle = e_3^* T(e_2) = \begin{pmatrix} \sqrt{3}/2 \\ 0 \\ 3\sqrt{3}/2 \end{pmatrix} = 0,
\]
which gives us the entries for the second column.

The entries for the third column of the $B$-matrix are obtained by computing $<T(e_3), e_1>$, $<T(e_3), e_2>$ and $<T(e_3), e_3>$, so proceeding as before, we have

$$T(e_3) = \begin{pmatrix} 0 & 1 & 3 \\ 1 & 1 & 1 \\ 3 & 1 & 0 \end{pmatrix} \begin{pmatrix} \sqrt{2} \\ \sqrt{2} \\ -\sqrt{2} \end{pmatrix} = \begin{pmatrix} -3\sqrt{2} \\ 0 \\ 3\sqrt{2} \end{pmatrix},$$

so

$$<T(e_3), e_1> = e_1^T T(e_3) = (0 \ 1 \ 0) \begin{pmatrix} -3\sqrt{2} \\ 0 \\ 3\sqrt{2} \end{pmatrix} = 0,$$

$$<T(e_3), e_2> = e_2^T T(e_3) = (\sqrt{2} \ 0 \ \sqrt{2}) \begin{pmatrix} -3\sqrt{2} \\ 0 \\ 3\sqrt{2} \end{pmatrix} = 0,$$

and

$$<T(e_3), e_3> = e_3^T T(e_3) = (\sqrt{2} \ 0 \ -\sqrt{2}) \begin{pmatrix} -3\sqrt{2} \\ 0 \\ 3\sqrt{2} \end{pmatrix} = -3,$$

and the $B$-matrix of $T$ is

$$\begin{pmatrix} 1 & \sqrt{2} \\ \sqrt{2} & 3 \ 0 \\ 0 & 0 \ -3 \end{pmatrix}_B.$$

**Problem 1** Assume $B = \{e_1, e_2, e_3\}$ with $e_1^t = (0, 0, 1)$, $e_2^t = (0, \sqrt{2}, 0, 0)$, and $e_3^t = (\sqrt{2}, 0, 0, 0)$. Give the Fourier expansion of the vector $u$ relative to $B$ when $u^t = (1, 0, 0)$, and if the standard matrix of $T$ is

$$\begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \\ 3 & 1 & 0 \end{pmatrix},$$

find the $B$-matix of $T$.

**Problem 2** Assume $B = \{e_1, e_2\}$ with $e_1^t = (\frac{1}{\sqrt{2}}, \frac{i}{\sqrt{2}})$ and $e_2^t = (\frac{i}{\sqrt{2}}, \frac{1}{\sqrt{2}})$. If the standard matrix of $T$ is

$$\begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix},$$

find the $B$-matix of $T$ by calculating the Fourier expansions of $T(e_1)$ and $T(e_2)$.

### 9.3 Graham-Schmidt

It is intuitively clear that a plane in $R^3$ must contain an orthonormal basis (in fact, infinitely many of them). We will now argue that any nontrivial subspace of $R^n$ contains an orthonormal basis. Consider all the orthonormal sets (sets whose elements have length one and are pairwise orthogonal) in the subspace. Among these sets are the singletons of the form $\{u\}$ with $<u, u> = 1$, so there exist such sets. Let $B$ be one of these sets with the largest number of elements. Since these sets are independent and inside of $R^n$, none of them can contain more than $n$ elements, so there certainly exists one of largest size, like our choosen one $B$. Let us enumerate the elements of $B$, like $B = \{e_1, \ldots e_k\}$. 


We will prove that $B$ must span the subspace, in which case it is an orthonormal basis. The proof proceeds by contradiction; the assumption that $B$ does not span the subspace leads to the conclusion that $B$ doesn’t have the largest number of elements among the orthonormal sets inside the subspace, contrary to the way it was chosen. If $B$ does not span the subspace, then there is a vector $u$ in the subspace that is not in the span of $B$, in which case

$$u \neq <u, e_1> e_1 + \ldots + <u, e_k> e_k,$$

and

$$v = u - <u, e_1> e_1 + \ldots + <u, e_k> e_k$$

is then a nonzero vector orthogonal to each $e_i$. If $e$ is the normalization of $v$, then

$${e_1, \ldots, e_k, e}$$

is an orthonormal set in the subspace with more elements than $B$.

This proof on the existence of an orthonormal basis suggests an algorithm that lets us construct an orthonormal set from a list of vectors, so that the orthonormal set has the same span as the list of vectors. The algorithm operates by repeatedly appealing to the above argument, which is a form of mathematical induction. Let’s illustrate by assuming we have a list of vectors $\{u_1, u_2, \ldots\}$. The first step in the algorithm is to look at $u_1$: if it is zero we do nothing, otherwise we let $e_1$ be its normalization. Then $\{e_1\}$ is an orthonormal set with the same span as $\{u_1\}$. The second step is to look at the span of $\{u_1, u_2\}$: if this is the same as the span of $\{e_1\}$ we do nothing, otherwise $u_2$ is not in the span of $\{e_1\}$ and we use the reasoning in the proof to say that $u_2 \neq <u_2, e_1> e_1$ and hence $v_2 = u_2 - <u_2, e_1> e_1 \neq 0$, and if $e_2$ is the normalization of $v_2$, then $\{e_1, e_2\}$ is an orthonormal set with the same span as $\{u_1, u_2\}$. If we have done this process $k$ times, we arrive at an orthonormal set $\{e_1, \ldots, e_j\}$ with the same span as $\{u_1, \ldots, u_k\}$. If $u_{k+1}$ is in this span we do nothing, otherwise we normalize

$$u_{k+1} - <u_{k+1}, e_1> e_1 + \ldots + <u_{k+1}, e_j> e_j$$

to obtain $e_{j+1}$, and $\{e_1, \ldots, e_{j+1}\}$ is an orthonormal set with the same span as $\{u_1, \ldots, u_{k+1}\}$. This process is called the Graham-Schmidt orthogonalization algorithm.

**Example 2**

Assume $M$ is the plane spanned by the vectors $u_1$ and $u_2$, where $u_1 = (1, 1, 0)$ and $u_2 = (1, 0, 1)$. Find an orthonormal basis of $M$.

We begin by letting $e_1$ be the normalization of $u_1$, so

$$e_1 = \frac{u_1}{\|u_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix}.$$

Now $u_2$ is not in the span of $e_1$, so we normalize $u_2 - <u_2, e_1> e_1$ to obtain $e_2$, like

$$u_2 - <u_2, e_1> e_1 = \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} - (1, 0, 1) \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \\ 0 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix},$$

$$\|u_2 - <u_2, e_1> e_1\| = \| \begin{pmatrix} \frac{1}{2} \\ -\frac{1}{2} \\ 1 \end{pmatrix} \| = \sqrt{\left(\frac{1}{2}\right)^2 + \left(-\frac{1}{2}\right)^2 + (1)^2} = \frac{\sqrt{6}}{2},$$
and

\[
e_2 = \frac{u_2 - \langle u_2, e_1 \rangle e_1}{\|u_2 - \langle u_2, e_1 \rangle e_1\|} = \frac{2}{\sqrt{6}} \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \\ 1 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{2}{\sqrt{6}} \end{pmatrix}.
\]

The set \( \{e_1, e_2\} \) is an orthonormal basis for the plane \( \mathcal{M} \).

**Problem 3** Assume \( \mathcal{M} \) is the plane spanned by the vectors \( u_1 \) and \( u_2 \), where \( u_1^t = (1, 1, 1) \) and \( u_2^t = (2, -1, 1) \). Find an orthonormal basis of \( \mathcal{M} \).

### 9.4 Tensors

The inner product on \( \mathbb{R}^n \) is an example of a function of two variables that is linear in both. Such functions are called **bilinear**, and there is a technique that reduces the study of bilinear functions to linear transformations. The idea is to embed the domain of the bilinear function inside a vector space, and construct a linear transformation on that space so that, when restricted to the embedded elements, the linear transformation coincides with the bilinear function. For example, the inner product on \( \mathbb{R}^3 \) is a function of two vector variables, so its domain is \( \mathbb{R}^3 \times \mathbb{R}^3 \), which is

\[
\{(u, v) : u, v \in \mathbb{R}^3\}.
\]

The vector space that this domain embeds into is called the **tensor product** and is denoted \( \mathbb{R}^3 \otimes \mathbb{R}^3 \). The embedding takes the pair \((u, v)\) to the element denoted \( u \otimes v \), and there is a linear transformation \( \Gamma : \mathbb{R}^3 \otimes \mathbb{R}^3 \to R \) so that, when applied to the embedded elements, the transformation \( \Gamma \) coincides with the inner product, in other words, \( \Gamma(u \otimes v) = \langle u, v \rangle \). It happens to be that \( \mathbb{R}^3 \otimes \mathbb{R}^3 \) is isomorphic to the \( 3 \times 3 \) matrices, with the trace playing the role of \( \Gamma \). The embedding is the one that takes a pair of vectors \( u \) and \( v \), given in their standard coordinates, to the matrix \( uv^* \), and we consequently let \( u \otimes v \) denote the transformation whose standard matrix is \( uv^* \). It follows that \( u \otimes v \) is a rank one transformation and

\[
u \otimes v(w) = \langle w, v \rangle u
\]

for all \( w \in \mathbb{R}^3 \). If \( e \) is a vector with length one, so that \( \langle e, e \rangle = 1 \), then

\[
(ee^*) (ee^*) = e (e^* e) e^* = e < e, e > e^* = ee^*;
\]

which says that \( (e \otimes e)^2 = e \otimes e \), so \( e \otimes e \) is a projection. It is, in fact, the orthogonal projection onto the line spanned by \( e \).

**Example 3** Find the standard matrix of the orthogonal projection onto the line spanned by \( u \) when

\[
u' = (1, 3, 2).
\]

We first normalize \( u \), computing \( \|u\| = \sqrt{1^2 + 3^2 + 2^2} = \sqrt{14} \), and dividing \( u \) by it length, getting

\[
e = \frac{u}{\|u\|} = \begin{pmatrix} \frac{1}{\sqrt{14}} \\ \frac{3}{\sqrt{14}} \\ \frac{2}{\sqrt{14}} \end{pmatrix}.
\]
The standard matrix for the orthogonal projection onto the line spanned by \( \mathbf{u} \) is simply \( \mathbf{e} \mathbf{e}^* \), which is
\[
\mathbf{e} \mathbf{e}^* = \begin{pmatrix}
\frac{1}{\sqrt{3}} \\
\frac{3}{\sqrt{14}} \\
\frac{2}{\sqrt{14}}
\end{pmatrix} \begin{pmatrix}
\frac{1}{\sqrt{14}} \\
\frac{3}{\sqrt{14}} \\
\frac{2}{\sqrt{14}}
\end{pmatrix} = \begin{pmatrix}
\frac{1}{14} \\
\frac{3}{14} \\
\frac{2}{14}
\end{pmatrix}.
\]

Compare this to the method used in example 4 on page 57.

The Fourier expansion of a vector
\[
\mathbf{u} = < \mathbf{u}, \mathbf{e}_1 > \mathbf{e}_1 + < \mathbf{u}, \mathbf{e}_2 > \mathbf{e}_2 + < \mathbf{u}, \mathbf{e}_3 > \mathbf{e}_3
\]
can be rewritten as
\[
I(\mathbf{u}) = (\mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3)(\mathbf{u}),
\]
and since this equality holds for every \( \mathbf{u} \in \mathbb{R}^3 \), we have
\[
I = \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3.
\]
The three transformations \( \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 + \mathbf{e}_3 \otimes \mathbf{e}_3 \), and \( \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_3 \otimes \mathbf{e}_3 \) are all orthogonal projections onto planes. Thus, when we have an orthonormal basis \( \mathcal{B} \), the orthogonal projections onto the various spans of elements of \( \mathcal{B} \) are obtainable as sums of the form \( \mathbf{e} \otimes \mathbf{e} \) with \( \mathbf{e} \in \mathcal{B} \). The rank of the projection will correspond to the length of the sum.

**Example 4**

Find the standard matrix of the orthogonal projection onto the plane spanned by \( \mathbf{u} \) and \( \mathbf{v} \) when
\[
\mathbf{u}^i = (-1, 2, 1) \quad \text{and} \quad \mathbf{v}^i = (1, 1, 1).
\]

We first use Gram-Schmidt on the vectors \( \mathbf{u} \) and \( \mathbf{v} \) to find an orthonormal basis for the plane. Let
\[
\mathbf{e}_1 = \frac{\mathbf{v}}{\|\mathbf{v}\|} = \begin{pmatrix}
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}}
\end{pmatrix},
\]
then remove from \( \mathbf{u} \) the part that lies in the direction of \( \mathbf{e}_1 \),
\[
\mathbf{u} - < \mathbf{u}, \mathbf{e}_1 > \mathbf{e}_1 = \begin{pmatrix}
-1 \\
2 \\
1
\end{pmatrix} - \frac{2}{\sqrt{3}} \begin{pmatrix}
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}}
\end{pmatrix} = \begin{pmatrix}
-\frac{5}{3} \\
\frac{4}{3} \\
\frac{1}{3}
\end{pmatrix},
\]
giving us a vector orthogonal to \( \mathbf{e}_1 \), which, after normalizing
\[
\mathbf{e}_2 = \frac{\mathbf{u} - < \mathbf{u}, \mathbf{e}_1 > \mathbf{e}_1}{\|\mathbf{u} - < \mathbf{u}, \mathbf{e}_1 > \mathbf{e}_1\|} = \frac{3}{\sqrt{42}} \begin{pmatrix}
-\frac{5}{3} \\
\frac{4}{3} \\
\frac{1}{3}
\end{pmatrix} = \begin{pmatrix}
-\frac{5}{\sqrt{42}} \\
\frac{4}{\sqrt{42}} \\
\frac{1}{\sqrt{42}}
\end{pmatrix},
\]
gives us our second basis vector. The projection onto the plane is now found by computing the standard matrix of \( \mathbf{e}_1 \otimes \mathbf{e}_1 + \mathbf{e}_2 \otimes \mathbf{e}_2 \), which is
\[
\mathbf{e}_1 \mathbf{e}_1^* + \mathbf{e}_2 \mathbf{e}_2^* = \begin{pmatrix}
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}}
\end{pmatrix} \begin{pmatrix}
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}} \\
\frac{1}{\sqrt{3}}
\end{pmatrix} + \begin{pmatrix}
-\frac{5}{\sqrt{42}} \\
\frac{1}{\sqrt{42}} \\
\frac{1}{\sqrt{42}}
\end{pmatrix} \begin{pmatrix}
-\frac{5}{\sqrt{42}} \\
\frac{4}{\sqrt{42}} \\
\frac{1}{\sqrt{42}}
\end{pmatrix}.
\]
After doing the arithmetic, we arrive at
\[
\begin{pmatrix}
\frac{39}{42} & -\frac{6}{42} & \frac{9}{42} \\
-\frac{6}{42} & \frac{30}{42} & \frac{18}{42} \\
\frac{9}{42} & \frac{18}{42} & \frac{15}{42}
\end{pmatrix}.
\]
We can check this answer using the fact that \( I - (e_1 \otimes e_1 + e_2 \otimes e_2) = e_3 \otimes e_3 \). It says that if we subtract our answer from the identity matrix, we should get the matrix for \( e_3 \otimes e_3 \), whose range should be orthogonal to both \( e_1 \) and \( e_2 \). We compute
\[
\begin{pmatrix}
1 & 0 & 0 \\
0 & 1 & 0 \\
0 & 0 & 1
\end{pmatrix} - \begin{pmatrix}
\frac{39}{42} & -\frac{6}{42} & \frac{9}{42} \\
-\frac{6}{42} & \frac{30}{42} & \frac{18}{42} \\
\frac{9}{42} & \frac{18}{42} & \frac{15}{42}
\end{pmatrix} = \begin{pmatrix}
\frac{3}{42} & \frac{6}{42} & -\frac{9}{42} \\
\frac{6}{42} & \frac{12}{42} & -\frac{18}{42} \\
-\frac{9}{42} & -\frac{18}{42} & \frac{27}{42}
\end{pmatrix},
\]
and see that each column is a multiple of the others, so it is indeed rank one, and each column is orthogonal to both \( e_1 \) and \( e_2 \).

**Problem 4** Find the standard matrix of the orthogonal projection onto the plane spanned by \( u \) and \( v \) when
\[
u^t = (3, 2, -1) \quad \text{and} \quad v^t = (-1, 1, 1).
\]
Verify your answer by subtracting it from the identity matrix.

### 9.5 Real and Imaginary Parts

There is an operation on transformations that is a generalization of conjugation for complex numbers. If \( T : C^n \rightarrow C^n \) is a linear transformation, then the adjoint of \( T \), denoted \( T^* \), is the unique linear transformation that satisfies
\[
< Tu, v > = < u, T^* v >
\]
for all vectors \( u, v \in C^n \). The adjoint can be defined in terms of standard matrices; if \( \{e_1, \ldots, e_n\} \) denotes the standard basis of \( C^n \), then the standard matrix of \( T \) has the entry \( < Te_j, e_i > \) in the \( i, j \)th position. The conjugate transpose of this matrix is obtained by taking the conjugate of every entry, and turning the \( i \)th row into the \( i \)th column for each \( i = 1, \ldots, n \). Thus, the conjugate transpose has the entry
\[
< Te_i, e_j > = < e_j, Te_i >
\]
in the \( i, j \)th position. If we define the adjoint of \( T \) to be the transformation with this standard matrix, then it follows that
\[
< e_j, Te_i > = < T^* e_j, e_i >
\]
for all of the standard basis vectors, and using the linearity of the inner product, we can extend this equality to
\[
< T^* u, v > = < u, T v >
\]
for all vectors \( u, v \in C^n \). In particular, if \( \{f_1, \ldots, f_n\} \) is another orthonormal basis, the we have
\[
< f_j, T^* f_i > = < T^* f_j, f_i >,
\]
which means that the matrices of \( T \) and \( T^* \) are conjugate transposes of each other relative to any orthonormal basis. If
\[
\begin{pmatrix}
a & b \\
c & d
\end{pmatrix} = \begin{pmatrix}
1 & 2 & 3 \\
4 & 5 & 6 \\
7 & 8 & 9
\end{pmatrix}, \quad \begin{pmatrix}
1 + i & 0 & 0 \\
2 - i & -1 & 0 \\
3 + 6 i & 1 & -i
\end{pmatrix}
\]
are the standard matrices of three transformations, then the respective standard matrices of their adjoints are

\[
\begin{pmatrix}
\bar{a} & \bar{c} \\
\bar{b} & \bar{d}
\end{pmatrix}
\begin{pmatrix}
1 & 4 & 7 \\
2 & 5 & 8 \\
3 & 6 & 9
\end{pmatrix}
\begin{pmatrix}
1 - i & 2 + i & 3 - 6i \\
0 & -1 & 1 \\
0 & 0 & i
\end{pmatrix}.
\]

A complex number has a real and imaginary part, letting us write \(z = a + ib\) with \(a\) and \(b\) real. If we know what \(z\) is, then we can find \(a\) and \(b\) with the formulae

\[
a = \frac{z + \bar{z}}{2} \quad \text{and} \quad b = \frac{z - \bar{z}}{2i}.
\]

By analogy, we can write any transformation as \(T = A + iB\) where \(A\) and \(B\) are found by

\[
A = \frac{T + T^*}{2} \quad \text{and} \quad B = \frac{T - T^*}{2i},
\]

and we call \(A\) and \(B\) the **real and imaginary parts** of \(T\) (respectively). Just as the numbers \(a\) and \(b\) are characterized as being real numbers by the fact that conjugation does nothing to them, the transformations \(A\) and \(B\) satisfy \(A = A^*\) and \(B = B^*\), and such transformations are called **Hermitian operators**.

**Example 5**

Decompose the following matrix into its real and imaginary parts:

\[
\begin{pmatrix}
1 - i & 2 + i & 3 - 6i \\
0 & -1 & 1 \\
0 & 0 & i
\end{pmatrix}.
\]

We compute

\[
A = \frac{\begin{pmatrix}
1 - i & 2 + i & 3 - 6i \\
0 & -1 & 1 \\
0 & 0 & i
\end{pmatrix} + \begin{pmatrix}
1 + i & 0 & 0 \\
2 - i & -1 & 0 \\
3 + 6i & 1 & -i
\end{pmatrix}}{2} = \begin{pmatrix}
1 & 1 + \frac{i}{2} & \frac{3}{2} - 3i \\
1 - \frac{i}{2} & -1 & \frac{1}{2} \\
\frac{3}{2} + 3i & \frac{1}{2} & 0
\end{pmatrix},
\]

and

\[
B = \frac{\begin{pmatrix}
1 - i & 2 + i & 3 - 6i \\
0 & -1 & 1 \\
0 & 0 & i
\end{pmatrix} - \begin{pmatrix}
1 + i & 0 & 0 \\
2 - i & -1 & 0 \\
3 + 6i & 1 & -i
\end{pmatrix}}{2i} = \begin{pmatrix}
-1 & \frac{1}{2} - i & -3 - \frac{3i}{2} \\
\frac{1}{2} + i & 0 & -\frac{i}{2} \\
-3 + \frac{3i}{2} & \frac{i}{2} & 1
\end{pmatrix},
\]

so we have the decomposition

\[
\begin{pmatrix}
1 - i & 2 + i & 3 - 6i \\
0 & -1 & 1 \\
0 & 0 & i
\end{pmatrix} = \begin{pmatrix}
1 & 1 + \frac{i}{2} & \frac{3}{2} - 3i \\
1 - \frac{i}{2} & -1 & \frac{1}{2} \\
\frac{3}{2} + 3i & \frac{1}{2} & 0
\end{pmatrix} + i \begin{pmatrix}
-1 & \frac{1}{2} - i & -3 - \frac{3i}{2} \\
\frac{1}{2} + i & 0 & -\frac{i}{2} \\
-3 + \frac{3i}{2} & \frac{i}{2} & 1
\end{pmatrix}.
\]

**Problem 5**

Decompose the following matrices into their real and imaginary parts:

i) \(\begin{pmatrix}
1 & 2 \\
3 & 4
\end{pmatrix}\)  
ii) \(\begin{pmatrix}
8 & -i \\
i & -1
\end{pmatrix}\)  
iii) \(\begin{pmatrix}
1 & 2i & 3 \\
0 & i & 4i \\
0 & 0 & 10
\end{pmatrix}\).
9.6 Spectral Theorem

The real and imaginary parts of a matrix are a generalization of the real and imaginary parts of a complex number, and the Hermitian matrices are the generalizations of real numbers. The spectral theorem helps make this correspondence more transparent; the theorem says that every Hermitian matrix is **orthogonally diagonalizable**, that is, it is diagonalizable using an orthonormal basis, and all of its eigenvalues that appear on the diagonal are real numbers. The proof of this theorem follows from the fact that every transformation $T : C^n \to C^n$ has a triangular matrix form relative to some orthonormal basis. When $T = T^*$, saying that $T$ is in triangular form relative to an orthonormal basis implies that it is actually diagonalized relative to that basis, and all the entries on the diagonal have to satisfy $\alpha = \bar{\alpha}$, so all the eigenvalues are real.

If $T$ is orthogonally diagonalizable, and $\mathcal{B} = \{f_1, \ldots, f_n\}$ is an orthonormal basis relative to which the $\mathcal{B}$-matrix is diagonal, then

$$T = \alpha_1 f_1 \otimes f_1 + \ldots + \alpha_n f_n \otimes f_n,$$

where $\alpha_1, \ldots, \alpha_n$ are the eigenvalues of $T$. When each $f_i$ is given in terms of its standard coordinates, the standard matrix of the above sum is then

$$\alpha_1 f_1 f_1^* + \ldots + \alpha_n f_n f_n^*,$$

and this is called the standard **spectral decomposition** of $T$.

**Example 6**

Find the standard spectral decomposition of the transformation whose standard matrix is

$$\begin{pmatrix}
1 & 2 \\
2 & 4
\end{pmatrix}.$$

Let $T$ denote the transformation with this standard matrix, and note that $T = T^*$, so the spectral theorem tells us that we can diagonalize $T$ with an orthonormal basis. When we invoke the diagonalization algorithm, we find that the characteristic polynomial is

$$x^2 - 5x,$$

which has roots $x = 5$ and $x = 0$. An eigenvector corresponding to $x = 5$ can be found in the nullspace of

$$\begin{pmatrix}
x - 1 & -2 \\
-2 & x - 4
\end{pmatrix}.$$

Searching for a dependency between the columns of

$$\begin{pmatrix}
4 & -2 \\
-2 & 1
\end{pmatrix},$$

reveals that the second column is $-\frac{1}{2}$ times the first column, and hence the vector

$$u_1^* = (-\frac{1}{2}, -1)$$

is an eigenvector of $T$. In a similar way, we find an eigenvector corresponding to $x = 0$, which is

$$u_2^* = (2, -1).$$
Notice that \( \mathbf{u}_1 \) is orthogonal to \( \mathbf{u}_2 \), so when we normalize these vectors we obtain the orthonormal basis
\[
\{ \mathbf{f}_1, \mathbf{f}_2 \} = \left\{ \begin{pmatrix} -\frac{1}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} \end{pmatrix}, \begin{pmatrix} \frac{2}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} \end{pmatrix} \right\},
\]
relative to which the matrix of \( T \) is diagonal. The standard spectral decomposition is then
\[
\alpha_1 \mathbf{f}_1 \mathbf{f}_1^* + \alpha_2 \mathbf{f}_2 \mathbf{f}_2^* = 5 \begin{pmatrix} -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ -\frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \end{pmatrix} + 0 \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \frac{2}{\sqrt{5}} & -\frac{1}{\sqrt{5}} \\ -\frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \end{pmatrix},
\]
which can be checked by comparing this sum to the matrix we started with:
\[
5 \begin{pmatrix} \frac{1}{5} & \frac{2}{5} \\ \frac{2}{5} & \frac{1}{5} \end{pmatrix} + 0 \begin{pmatrix} \frac{4}{5} & -\frac{2}{5} \\ -\frac{2}{5} & \frac{4}{5} \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}.
\]

**Problem 6** Find the standard spectral decompositions of the transformations whose standard matrices are
\[
i) \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \quad ii) \begin{pmatrix} 8 & -i \\ i & -1 \end{pmatrix} \quad iii) \begin{pmatrix} 1 & 2i & 0 \\ -2i & 1 & -i \\ 0 & i & 1 \end{pmatrix}.
\]

A spectral decomposition can, of course, be given relative to any orthonormal basis \( \mathcal{B} \). If we write \( \mathbf{f}_1 \) and \( \mathbf{f}_2 \) from example 6 in terms of their \( \mathcal{B} \)-coordinates, then the \( \mathcal{B} \)-spectral decomposition is
\[
\alpha_1 \mathbf{f}_1 \mathbf{f}_1^* + \alpha_2 \mathbf{f}_2 \mathbf{f}_2^*,
\]
where all the computations are carried out using the \( \mathcal{B} \)-coordinates instead of the standard coordinates, and the sum should add up to the \( \mathcal{B} \)-matrix of \( T \). The simplest of these decompositions happens when \( \mathcal{B} = \{ \mathbf{f}_1, \mathbf{f}_2 \} \), in which case we get
\[
\begin{pmatrix} 5 & 0 \\ 0 & 0 \end{pmatrix}_\mathcal{B} = 5 \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_\mathcal{B} + 0 \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}_\mathcal{B}.
\]

There are a few transformations that are orthogonally diagonalizable even though they are not Hermitian. If \( T \) satisfies \( TT^* = T^*T \), then \( T \) is called a normal operator, and this implies that \( AB = BA \) when \( A \) and \( B \) denote the real and imaginary parts of \( T \). Two commuting transformations, such as \( A \) and \( B \), can be simultaneously triangularized, that is, they can be put into triangular form using the same orthonormal basis. When this is done, both \( A \) and \( B \) will actually be diagonalized (since they are both Hermitian), and hence \( T \) will itself be diagonalized, since \( T = A + iB \). It is even more apparent that any transformation that can be orthogonally diagonalized must be a normal operator, thus a more general form of the spectral theorem states that a transformation is orthogonally diagonalizable if and only if it is a normal operator.

**Example 7** Find out which of the following matrices are orthogonally diagonalizable:
\[
i) \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} \quad ii) \begin{pmatrix} i & -i \\ i & -1 \end{pmatrix} \quad iii) \begin{pmatrix} 0 & -3 & 5 \\ 3 & 0 & 4 \\ -5 & -4 & 0 \end{pmatrix} \quad iv) \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.
\]
If $T$ denotes the first matrix, then

$$TT^* = \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix} = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} -i & 1 \\ 1 & -i \end{pmatrix} \begin{pmatrix} i & 1 \\ 1 & i \end{pmatrix} = T^*T,$$

so $T$ is normal, and hence orthogonally diagonalizable. By contrast, when $T$ is the second matrix, we get

$$TT^* = \begin{pmatrix} i & -i \\ i & -1 \end{pmatrix} \begin{pmatrix} -i & -i \\ i & -1 \end{pmatrix} = \begin{pmatrix} 2 & 1+i \\ 1-i & 2 \end{pmatrix},$$

which is not the same as

$$\begin{pmatrix} 2 & i-1 \\ -1-i & 2 \end{pmatrix} = \begin{pmatrix} -i & -i \\ i & -1 \end{pmatrix} \begin{pmatrix} i & -i \\ -i & 1 \end{pmatrix} = T^*T,$$

so this matrix is not orthogonally diagonalizable.

We could compute $TT^*$ and compare it with $T^*T$ for the third matrix, but here is a faster way to do it; notice that $T^* = -T$, so $T^*$ commutes with $T$, and $T$ is normal, hence orthogonally diagonalizable. The last matrix is Hermitian, so it is also normal and orthogonally diagonalizable.

**Problem 7** Find out which of the following matrices are orthogonally diagonalizable;

i) \( \begin{pmatrix} i & -1 \\ 1 & i \end{pmatrix} \) ii) \( \begin{pmatrix} 1+i & 1 \\ 1 & 1-i \end{pmatrix} \) iii) \( \begin{pmatrix} 0 & i & \tilde{i} \\ i & 0 & i \\ \tilde{i} & i & 0 \end{pmatrix} \) iv) \( \begin{pmatrix} 0 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \).

### 9.7 Polar Decomposition

A complex number $z$ can be factored as a product of the form $ur$, where $r$ is a positive real number that equals the length of $z$, and $u$ is a complex number on the unit circle, so $u$ has length one and determines the direction of $z$. This decomposition generalizes to matrices, resulting in a factorization of a matrix into what is called its **polar decomposition**. If $T$ is a linear transformation on $\mathbb{C}^n$, then $T$ can be factored $T = UP$, where $P$ is the matrix generalization of a positive real number, called a **positive operator**, and $U$ is the generalization of a complex number of length one.

The generalization of a positive number is an orthogonally diagonalizable matrix which, in its diagonal form, has positive numbers on the diagonal. These transformations are called **positive operators**, and the matrices themselves are called **positive matrices**. Some people distinguish between the invertible and noninvertible positive matrices, using the phrases **positive definite** and **positive semidefinite**, respectively. Given any transformation $T$, we can form a positive operator $T^*T$, which in its diagonal form, has positive numbers down the diagonal. If we take the square roots of all these diagonal entries, we obtain another positive matrix, whose corresponding operator is denoted $(T^*T)^{\frac{1}{2}}$ or $|T|$. If $T$ were positive to begin with, then $T = |T|$. We refer to $|T|$ as the **positive part** of $T$.

When we consider the linear transformations on a one dimensional space, there really are not too many; they are all scalar transformations, functions that take a vector $u$ to $zu$, for some scalar $z$. Among the scalars $z$, the ones with length one are exactly those that determine **isometries**, which are functions that preserve distance. The matrix generalization is called a **unitary matrix**, with the corresponding transformation called a **unitary operator**. Such a transformation $U$ is defined to satisfy

$$||Uu|| = ||u||$$

for every $u$, which says that the length of $u$ should equal the length of its image under $U$. The linearity lets one deduce that a unitary actually preserves distances and angles, as well. Invertible
transformations are exactly the ones that take every basis to another basis, and unitary operators are the transformations that take every orthonormal basis to another orthonormal basis. If we write the matrix of a unitary relative to an orthonormal basis, we should see that the columns are mutually orthogonal and all of length one.

A relationship between a transformation $T$ and its positive part $|T|$ begs for the construction of a unitary, the relationship being

$$||Tu|| = |||Tu||.$$

The unitary takes $|Tu|$ to $Tu$, from which we automatically get $Tu = U|Tu|$ and $T = U|T|$.

**Example 8**

Find the positive part of the matrix

$$\begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix}.$$

We will consider this matrix to be the standard matrix of $T$, so that the standard matrix of $T^*T$ is

$$T^*T = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

The next step is to diagonalize this matrix, so we can compute the square root. The diagonalization algorithm will lead us to a basis like

$$B = \{f_1, f_2\} = \left\{ \begin{pmatrix} \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix}, \begin{pmatrix} -\frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} \end{pmatrix} \right\},$$

where the $B$-matrix of $T^*T$ is

$$\begin{pmatrix} 2 & 0 \\ 0 & 0 \end{pmatrix}_B,$$

so the $B$-matrix of $(T^*T)^{\frac{1}{2}}$ is

$$\begin{pmatrix} \sqrt{2} & 0 \\ 0 & 0 \end{pmatrix}_B.$$

Returning to the standard coordinate system, we note that finding the standard matrix of $(T^*T)^{\frac{1}{2}}$ is simplified by the observation that, in this particular case, $(T^*T)^{\frac{1}{2}} = \frac{\sqrt{2}}{2} T^*T$, so the standard matrix of $(T^*T)^{\frac{1}{2}}$ is

$$\frac{\sqrt{2}}{2} \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \\ \frac{\sqrt{2}}{2} & \frac{\sqrt{2}}{2} \end{pmatrix},$$

which gives us the positive part.

**Problem 8**

Find the positive part of the matrix

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

**Example 9**

Find the polar decomposition of the matrix

$$\begin{pmatrix} 3 & \sqrt{2} \\ 0 & 2 \end{pmatrix}.$$
Let $T$ denote the transformation with this standard matrix, so the standard matrix of $T^* T$ is

$$T^* T = \begin{pmatrix} 3 & 0 \\ \sqrt{2} & 2 \end{pmatrix} \begin{pmatrix} 3 & \sqrt{2} \\ 0 & 2 \end{pmatrix} = \begin{pmatrix} 9 & 3\sqrt{2} \\ 3\sqrt{2} & 6 \end{pmatrix}.$$  

The characteristic polynomial of $T^* T$ is

$$(x - 9)(x - 6) - 18 = x^2 - 15x + 36 = (x - 12)(x - 3),$$

so the eigenvalues of $T^* T$ are $x = 12$ and $x = 3$. Plugging $x = 12$ into

$$\begin{pmatrix} x - 9 & -3\sqrt{2} \\ -3\sqrt{2} & x - 6 \end{pmatrix}$$

gives us

$$\begin{pmatrix} 3 & -3\sqrt{2} \\ -3\sqrt{2} & 6 \end{pmatrix},$$

and $-\sqrt{2}$ times the first column equals the second column, which leads us to the vector $(\sqrt{2}, 1)$ in the nullspace. If we normalize this vector, we obtain

$$f_1 = \begin{pmatrix} \frac{\sqrt{2}}{\sqrt{3}} \\ \frac{1}{\sqrt{3}} \end{pmatrix},$$

which is the first vector in an orthonormal basis of eigenvectors for $T^* T$. Repeating the process with the second eigenvalue $x = 3$ gives us the matrix

$$\begin{pmatrix} -6 & -3\sqrt{2} \\ -3\sqrt{2} & -3 \end{pmatrix},$$

which has $(\frac{1}{\sqrt{2}}, -1)$ in the nullspace, and normalizing this vector gives us

$$f_2 = \begin{pmatrix} \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} \end{pmatrix},$$

and an orthonormal basis of eigenvectors $B = \{ f_1, f_2 \}$, relative to which, the $B$-matrix of $T^* T$ is

$$\begin{pmatrix} 12 & 0 \\ 0 & 3 \end{pmatrix}_B.$$

Now we are able to see a matrix form for $(T^* T)^{\frac{1}{2}}$, namely

$$\begin{pmatrix} \sqrt{12} & 0 \\ 0 & \sqrt{3} \end{pmatrix}_B.$$

The standard matrix for $(T^* T)^{\frac{1}{2}}$ is then

$$\sqrt{12} f_1 f_1^* + \sqrt{3} f_2 f_2^* = \sqrt{12} \left( \begin{pmatrix} 2 \\ \sqrt{2} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{3} \\ 1 \end{pmatrix} + \sqrt{3} \begin{pmatrix} \frac{1}{3} \\ -\frac{\sqrt{2}}{3} \end{pmatrix} \right) = \begin{pmatrix} \frac{5\sqrt{3}}{3} \\ \frac{\sqrt{2}}{3} \end{pmatrix} \begin{pmatrix} \frac{\sqrt{2}}{3} \\ \frac{\sqrt{2}}{3} \end{pmatrix},$$

which is the standard matrix of the positive part $|T|$ of the polar decomposition.

The unitary part $U$ takes $|T|u$ to $T u$ for every $u$, so taking $u$ to be $f_1$ and $f_2$ we see that $\sqrt{12} \left( \begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix}, \begin{pmatrix} \sqrt{2} \\ 1 \end{pmatrix} \right)$ is taken to $(\frac{\sqrt{2}}{\sqrt{3}}, \frac{2}{\sqrt{3}})$ and $\sqrt{3} \left( \begin{pmatrix} \frac{\sqrt{2}}{3} \\ \frac{1}{3} \end{pmatrix}, \begin{pmatrix} -\frac{\sqrt{2}}{3} \\ \frac{1}{3} \end{pmatrix} \right)$ is taken to $(-\frac{1}{\sqrt{3}}, -\frac{\sqrt{2}}{\sqrt{3}})$, which is enough
information to determine what the standard matrix of \( U \) is. An alternative way of finding the standard
matrix of \( U \) is to find the standard matrix of \(|T|^{-1}\), then compute the product \( T|T|^{-1} \). This works
because of the relation \( T = U|T| \), which is equivalent to \( T|T|^{-1} = U \). Proceeding in this direction,
we note that the \( B \)-matrix of \(|T|^{-1} \) is
\[
\begin{pmatrix}
\frac{1}{\sqrt{12}} & 0 \\
0 & \frac{1}{\sqrt{3}}
\end{pmatrix}_B,
\]
so the standard matrix of \(|T|^{-1} \) is
\[
\frac{1}{\sqrt{12}} f_1 f_1^* + \frac{1}{\sqrt{3}} f_2 f_2^* = \frac{1}{\sqrt{12}} \begin{pmatrix} \frac{2}{3} & \frac{\sqrt{2}}{3} \\
\frac{\sqrt{2}}{3} & \frac{1}{3} \end{pmatrix} + \frac{1}{\sqrt{3}} \begin{pmatrix} \frac{1}{3} & -\frac{\sqrt{2}}{3} \\
-\frac{\sqrt{2}}{3} & \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3\sqrt{3}} & -\frac{\sqrt{2}}{6\sqrt{3}} \\
-\frac{\sqrt{2}}{3\sqrt{3}} & \frac{5}{6\sqrt{3}} \end{pmatrix},
\]
and
\[
U = T|T|^{-1} = \begin{pmatrix} 3 & \sqrt{2} \\
0 & 2 \end{pmatrix} \begin{pmatrix} \frac{2}{3\sqrt{3}} & -\frac{\sqrt{2}}{6\sqrt{3}} \\
\frac{\sqrt{2}}{6\sqrt{3}} & \frac{5}{6\sqrt{3}} \end{pmatrix} = \begin{pmatrix} \frac{5}{3\sqrt{3}} & \frac{\sqrt{2}}{3\sqrt{3}} \\
-\frac{\sqrt{2}}{3\sqrt{3}} & \frac{5}{3\sqrt{3}} \end{pmatrix}.
\]
A quick check of the columns shows that they are mutually orthogonal and of length one, so \( U \) is
indeed a unitary, and we are probably mistake free. Thus, the polar decomposition of \( T \), in standard
matrix form, looks like
\[
\begin{pmatrix} 3 & \sqrt{2} \\
0 & 2 \end{pmatrix} = \begin{pmatrix} \frac{5}{3\sqrt{3}} & \frac{\sqrt{2}}{3\sqrt{3}} \\
\frac{\sqrt{2}}{3\sqrt{3}} & \frac{5}{3\sqrt{3}} \end{pmatrix} \begin{pmatrix} \frac{5\sqrt{3}}{3} & \frac{\sqrt{6}}{3} \\
\frac{\sqrt{6}}{3} & 4\sqrt{3} \end{pmatrix}.
\]

**Problem 9** Find the polar decomposition of the matrix
\[
\begin{pmatrix} -2 & \sqrt{2} \\
0 & 3 \end{pmatrix}.
\]

### 9.8 Singular Value Decomposition

The polar decomposition is a special case of a slightly more general entity called the *singular value decomposition*. If \( T \) is a square complex matrix, then \(|P|\) is orthogonally diagonalizable, so
\(|P| = VDV^* \) for some diagonal matrix \( D \) and unitary matrix \( V \), and then the polar decomposition becomes
\[
T = U(VDV^*) = (UV)DV^* = WDV^*,
\]
where \( W = UV \) is also unitary. This is the singular value decomposition of \( T \); \( T = WDV^* \), with
\( V \) and \( W \) unitary matrices and \( D \) a diagonal matrix. This decomposition can be generalized to
arbitrary \( m \times n \) complex matrices \( T \). The assertion is that an \( m \times n \) complex matrices \( T \) can be
factored as \( T = WDV^* \) with \( V \) an \( n \times n \) unitary, \( D \) an \( m \times n \) diagonal matrix, and \( W \) an \( m \times m \)
unitary. One strategy for finding this decomposition runs as follows. Assume you are given an \( m \times n \)
complex matrix \( T \). Work either with \( T \) or with \( T^* \), so that on your working matrix the number of rows
exceeds the number of columns (for notational reasons, let us assume that \( T \) is the matrix to
work with, i.e. that \( m > n \)). Create an \( m \times m \) square matrix \( \hat{T} \) by inserting the appropriate number
of zero columns to the right of \( T \). Compute the polar decomposition of \( T \), getting \( \hat{T} = U|\hat{T}| \) for
some \( m \times m \) unitary \( U \), and then find an \( m \times m \) unitary \( \hat{V} \) and an \( m \times m \) diagonal matrix \( \hat{D} \)with
\(|\hat{T}| = \hat{V}\hat{D}V^* \). You then have

### 9.9 Discrete Fourier Transform
When we digitally record our voices while speaking into a microphone, a sequence of numbers is obtained that represent air pressure at different moments in time. The sequence of numbers is finite, but it is likely to be quite large. We treat the sequence as one vector in a high dimensional space, with the recorded numbers being the standard coordinates, and the standard coordinate system is referred to as the *time domain*. There is another coordinate system that is intimately related to periodic functions, and when we convert our vector into this coordinate system, the new coordinates are thought of as comprising the *fundamental harmonics* of our recorded sound. This coordinate system is referred to as the *frequency domain*, and its orthonormal basis is obtained as follows: if our recorded sound has $n$ components, we find a complex number $\alpha$ so that $\alpha^n = 1$, but $\alpha^k \neq 1$ for every $k < n$. Such a number is called a *primitive $n^{th}$ root of unity*. If we then, for $k = 0, 1, 2, \ldots n - 1$, form the vectors

$$v_k^t = (1, \alpha^k, \alpha^{2k}, \ldots, \alpha^{(n-1)k}),$$

then, as we will see, we obtain an orthogonal set of vectors, which upon normalization

$$f_k = \frac{v_k}{||v_k||},$$

results in our orthonormal basis. The essence of signal processing is to convert the vector from the time domain to the frequency domain, study the behavior of the harmonics, modify the harmonics to achieve a desired result, then convert back to the time domain and listen to the new recorded signal. The simplest thing to do is to zero out coordinates in the frequency domain that are smaller than a set threshold. This often has the effect of removing random noise from the original time domain signal. If a great many zeros are introduced, one may also encode the modified signal in a compressed form.

**Example 10**

Convert the vector $(1, -1, 0)$ into the frequency domain, compress the vector, then convert back to the time domain.

There are three coordinates, so we are imagining a very small sound sample as a vector in $R^3$. We then look for a *primitive cube root of identity*, i.e. a complex number $\alpha$ such that $\alpha^3 = 1$, but no smaller power of $\alpha$ equals one. A way to visualize how to get this number is to imagine the unit cube divided into three equal parts, with one of the divisions being the positive real axis. The three corresponding points on the unit circle are $1, -1 + \frac{\sqrt{3}}{2}i, \text{ and } -\frac{1}{2} - \frac{\sqrt{3}}{2}i$. Two of these numbers are primitive cube roots, and either one will serve as $\alpha$. We will choose

$$\alpha = -\frac{1}{2} + \frac{\sqrt{3}}{2}i.$$

The transposes of the three resulting vectors, $(1, \alpha^k, \alpha^{2k})$ with $k = 0, 1, 2$, are then

$$(1, 1, 1), \ (1, -\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} + \frac{\sqrt{3}}{2}i), \text{ and } (1, -\frac{1}{2} - \frac{\sqrt{3}}{2}i, -\frac{1}{2} + \frac{\sqrt{3}}{2}i).$$

**Problem 10**

Verify that these three vectors are mutually orthogonal.

Upon normalization, we obtain the orthonormal basis

$$f_1^t = \frac{1}{\sqrt{n}}(1, 1, 1), \ f_2^t = \frac{1}{\sqrt{n}}(1, -\frac{1}{2} + \frac{\sqrt{3}}{2}i, -\frac{1}{2} - \frac{\sqrt{3}}{2}i), \text{ and } f_3^t = \frac{1}{\sqrt{n}}(1, -\frac{1}{2} - \frac{\sqrt{3}}{2}i, -\frac{1}{2} + \frac{\sqrt{3}}{2}i).$$

Letting $v$ denote the transpose of $(1, -1, 0)$, we see that the coordinates in the frequency domain are $<v, f_1>$, $<v, f_2>$, and $<v, f_3>$, which are (respectively)
Chapter 10

Operator Theory

10.1 Hilbert Space

Spaces that occur in the mathematical formulations of quantum mechanics tend to be infinite dimensional versions of the complex inner product spaces discussed in the previous chapter. The simplest example is the space $l_2$ of all sequences of complex numbers $(u_i)$ with

$$\sum_{i=1}^{\infty} |u_i|^2 < \infty.$$ 

The theory is traditionally developed abstractly, by assuming one has a complex vector space $H$ with an inner product $\langle \cdot, \cdot \rangle$, which gives rise to a norm

$$||u|| = \sqrt{\langle u, u \rangle},$$

and this norm then gives rise to a metric, which defines the distance between $u$ and $v$ to be $||u - v||$. The space $H$ is called a **Hilbert Space** when this metric is complete. The sense of isomorphism between two Hilbert spaces $H$ and $K$ is determined by the existence of a unitary transformation

$$U : H \to K,$$

which is a bijection that preserves the inner product, i.e.

$$\langle u, v \rangle = \langle U(u), U(v) \rangle,$$

for all $u, v \in H$. Two isomorphic Hilbert spaces are structurally identical, and a fundamental fact is that each isomorphism class is determined by its dimension: two Hilbert spaces are isomorphic if and only if they have the same dimension, i.e. each contains an orthonormal basis and these two bases have the same cardinality. While the Hilbert spaces that arise in applications are typically spaces of functions, they are all isomorphic to $l_2$ and its subspaces, making $l_2$ a model for Hilbert space.
Formalism
Chapter 11

Bases

11.1 Vector Spaces

1. Definition of a vector space. A **Vector Space** $\mathcal{V}$ is a set, whose elements are called **vectors**, that has all of the following properties. There is a notion of addition on $\mathcal{V}$ so that $v, w \in \mathcal{V}$ implies $v + w \in \mathcal{V}$. The set $\mathcal{V}$ contains an additive identity element $\mathbf{0}$ that we call the zero vector, and $v \in \mathcal{V}$ implies $\mathbf{0} + v = v$. The addition is assumed to be associative, commutative, and to every vector $v \in \mathcal{V}$ there corresponds an inverse vector $-v \in \mathcal{V}$, so that $v + (-v) = \mathbf{0}$. There is also a notion of scalar multiplication defined with the scalars constituting the elements of a scalar field. If $\alpha$ and $\beta$ denote scalars, and $v, w \in \mathcal{V}$, we require that $\alpha(v + w) = \alpha v + \alpha w$, $(\alpha + \beta)v = \alpha v + \beta v$, and $(\alpha\beta)v = \alpha(\beta v)$, where all of the sums and products give another element of $\mathcal{V}$. Finally, when 1 denotes the multiplicative identity in the scalar field, we demand that $1v = v$ for every $v \in \mathcal{V}$. This mouthful is exactly what is needed to manipulate the vectors in $\mathcal{V}$ just as if they were vectors in $\mathbb{R}^n$ or $\mathbb{C}^n$.

If $\mathcal{V}$ is a vector space and $\mathcal{L} \subseteq \mathcal{V}$, then the addition and scalar multiplication on $\mathcal{V}$ also act on the subset $\mathcal{L}$. If $\mathcal{L}$ is itself a vector space with these operations we will call $\mathcal{L}$ a **subspace** of $\mathcal{V}$.

**Problem 1** Assume that $\mathcal{L} \subseteq \mathcal{V}$ is nonempty. Show that $\mathcal{L}$ is a subspace of $\mathcal{V}$ if and only if $\alpha v + \beta w \in \mathcal{L}$ whenever $\alpha$ and $\beta$ are scalars and $v, w \in \mathcal{L}$.

2. Linear combinations and span of a set. An expression of the form $\alpha v + \beta w$ is called a **linear combination** of $v$ and $w$. More generally, if $\mathcal{S}$ is a subset of a vector space $\mathcal{V}$, then a **linear combination** of vectors in $\mathcal{S}$ is an expression of the form

$$\alpha_1 v_1 + \ldots + \alpha_n v_n$$

with $v_1, \ldots, v_n \in \mathcal{S}$ and $\alpha_1, \ldots, \alpha_n$ scalars. Note that linear combinations are finite sums, while the set $\mathcal{S}$ may be a finite set or an infinite set.

The span of the set $\mathcal{S}$, denoted $\text{span}(\mathcal{S})$, is defined to be the set of all linear combinations of vectors in $\mathcal{S}$.

**Problem 2** Assume that $\mathcal{S} \subseteq \mathcal{V}$. Show that $\text{span}(\mathcal{S})$ is the smallest subspace of $\mathcal{V}$ containing $\mathcal{S}$.

3. Linear independence and spanning sets. A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is called a **spanning set** if $\text{span}(\mathcal{S}) = \mathcal{V}$, and in this case one occasionally abbreviates this by saying that $\mathcal{S}$ spans $\mathcal{V}$. A subset $\mathcal{S}$ of a vector space $\mathcal{V}$ is linearly independent when $v_1, \ldots, v_n \in \mathcal{S}$ and $\alpha_1 v_1 + \ldots + \alpha_n v_n = \mathbf{0}$ implies $\alpha_1 = \ldots = \alpha_n = 0$. If $\mathcal{S}$ is not independent we say it is **dependent**. If $v = \sum_i \alpha_i s_i$ is a non-zero vector in the span of $\mathcal{S}$, with each $s_i \in \mathcal{S}$, then at least one of the scalars $\alpha_i$ must be non-zero. If we were to add the vector $v$ to $\mathcal{S}$ and remove $s_j$ from $\mathcal{S}$ ($s_j$ is the vector being multiplied by the nonzero scalar $\alpha_j$), the new set will have the same span as $\mathcal{S}$. This fact can be used to prove that finite independent sets have fewer vectors than spanning sets.
Problem 3 Assume that $S_1$ is an independent subset of $V$ with $n$ elements and $S_2$ is a spanning set in $V$. Prove that $S_2$ has at least $n$ elements.

4. Bases and dimension. A subset $S \subseteq V$ is called a basis of $V$ if $S$ spans $V$ and $S$ is linearly independent. Two bases of $V$ will always have the same number of elements, and this number is called the dimension of the vector space. When both the bases are infinite, the proof of this assertion requires a fair bit of set theory to develop, but when the bases are finite the proof can be obtained as a corollary of Problem 3.

Problem 4 Assume $V$ has a finite spanning set, and let $B_1$ and $B_2$ denote two bases of $V$. Show that $B_1$ has the same number of elements as $B_2$.

5. Existence of bases. Every vector space $V$ contains a basis. The proof of this assertion is much easier in the case when $V$ has a finite spanning set. When no such spanning set exists, the proof requires some knowledge of the axiom of choice, but when finite spanning sets are present, one needs only select an independent set with the most number of elements.

Problem 5 Prove that a vector space with a finite spanning set contains a basis.

6. Extending to a basis. The same reasoning used in the previous problem will actually prove a stronger statement: given an independent subset $X$ of a vector space, there exists a basis of the vector space that contains $X$. The process of finding a basis that contains $X$ is called extending $X$ to a basis. Instead of blindly choosing an independent subset with the most number of elements, as we did to show that a basis exists, we select among the independent sets that contain $X$ to prove the more general assertion. To ensure the existence of an independent set ‘with the most elements’, we need the hypothesis that there is a finite spanning set.

Problem 6 Assume $X$ is an independent subset of a finite dimensional vector space. Show that there is a basis of the vector space that contains $X$.

7. Complements. If $M$ is a subspace of a vector space $V$, then $M$ has a basis $X$. The set $X$, being an independent subset of $V$, can be extended to a basis $B$ of $V$. Thus the basis $B$ can be expressed as a disjoint union $B = X \cup Y$, and if $N$ is the span of $Y$, then $N$ is also a subspace of $V$ and we say that $N$ is a complement of $M$ in $V$.

Problem 7 Show that $N$ is a complement of $M$ in $V$ if and only if $M \cap N = \{0\}$ and $\text{span}(M \cup N) = V$.

8. Rank and nullity. A linear transformation is a function $T: V \rightarrow W$ whose domain and codomain are vector spaces, and which satisfies

$$T(\alpha v + \beta w) = \alpha T(v) + \beta T(w)$$

for every linear combination in the domain. A linear transformation gives rise to two important subspaces. The set of vectors that are taken to the zero vector is called the nullspace of $T$. The nullspace is a subspace of the domain and its dimension is called the nullity of $T$. The range of the transformation $T$ is a subspace of its codomain, and the dimension of this range is called the rank of $T$. The rank and nullity theorem says that the rank plus the nullity equals the dimension of the domain. This assertion may be established by building a basis of the domain, beginning with a basis $X$ for the nullspace, then extending this to a basis $B = X \cup Y$ for the domain. The dimension of the nullspace, which is the number of elements in $X$, plus the number of elements in $Y$, equals the dimension of the domain. The final step is to show that

$$\{T(u) : u \in Y\}$$
is a basis for the range of $T$.

**Problem 8** If $T : \mathcal{V} \to \mathcal{W}$ is a linear transformation, show that the sum of the rank and the nullity of $T$ equals the dimension of $\mathcal{V}$.

**More Exercises**

1. The axioms of a vector space are stated in the first paragraph of chapter 11. With $\mathbf{u}$ and $\mathbf{v}$ denoting vectors in a vector space, use the axioms to prove the following statements.

   a. $\mathbf{0} + \mathbf{o} = \mathbf{o}$  
   b. $2\mathbf{v} = \mathbf{v} + \mathbf{v}$  
   c. $\mathbf{v} + \mathbf{u} = \mathbf{o}$  
   d. $(-1)\mathbf{u} = -\mathbf{u}$

2. Give an example of a subset $\mathcal{L}$ of $\mathbb{R}^2$ that satisfies $\alpha \mathbf{v} \in \mathcal{L}$, for all $\alpha \in \mathbb{R}$ and all $\mathbf{v} \in \mathcal{L}$, but such that $\mathcal{L}$ is not a subspace of $\mathbb{R}^2$.

3. Assume $\mathcal{L}$ is a subspace of $\mathcal{V}$. Show that for every natural number $n \geq 2$, scalars $\alpha_1, \ldots, \alpha_n$ and vectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathcal{L}$ we have $\alpha_1 \mathbf{v}_1 + \ldots + \alpha_n \mathbf{v}_n \in \mathcal{L}$.

4. Show that a subspace is equal to its span.

5. Show that intersections of spanning sets in $\mathcal{V}$ need not span $\mathcal{V}$.

6. Show that unions of spanning sets are spanning sets.

7. Assume that $\mathcal{S}$ is an independent set. Show that every vector in $\text{span}(\mathcal{S})$ has a unique representation as $\sum \alpha_i \mathbf{v}_i$ with each $\mathbf{v}_i \in \mathcal{S}$.

8. Show that intersections of independent sets in $\mathcal{V}$ are independent.

9. Show that unions of independent sets need not be independent.

10. Assume that $\mathcal{S}_1 \subseteq \mathcal{S}_2 \subseteq \mathcal{S}_3 \subseteq \ldots$ are independent subsets of $\mathcal{V}$. Show that $\cup_{i=1}^{\infty} \mathcal{S}_i$ is independent.

11. Assume $\mathbf{v} = \sum \alpha_i \mathbf{s}_i$ is a non-zero vector in the span of $\mathcal{S}$, with $\mathbf{s}_i \in \mathcal{S}$ for each $i$, and $\alpha_j$ a nonzero scalar for some value of $j$. Show that the set obtained by adding $\mathbf{v}$ to $\mathcal{S}$, then removing the vector $\mathbf{s}_j$ (i.e. removing the vector being multiplied by the nonzero scalar $\alpha_j$) has the same span as $\mathcal{S}$.

12. Assume that $\mathcal{V}$ has a finite basis. Show that any other basis of $\mathcal{V}$ must also be finite.

13. Show that two finite bases of $\mathcal{V}$ have the same number of elements.

14. Show that any finite spanning set contains a subset that is a basis.

15. Show that the nullspace of a transformation is a subspace of the transformation’s domain.

16. Show that the range of a transformation is a subspace of the transformation’s codomain.

17. Prove that a linear transformation is one-to-one if and only if the nullspace of $T$ consists of a single vector.

18. Assume that $T : \mathcal{V} \to \mathcal{V}$ is linear and $\mathcal{V}$ is finite dimensional. Prove that $T$ is one-to-one if and only if it is onto.

19. Give an example of a linear transformation $T : \mathcal{V} \to \mathcal{V}$ that is one-to-one, but not onto.

20. Give an example of a linear transformation $T : \mathcal{V} \to \mathcal{V}$ that is onto, but not one-to-one.

21. Assume that $\mathcal{V}$ is a vector space, and $\mathcal{M}$ is a complement of $\mathcal{N}$ in $\mathcal{V}$. Show that every vector $\mathbf{v} \in \mathcal{V}$ can be written uniquely as $\mathbf{v} = \mathbf{m} + \mathbf{n}$, with $\mathbf{m} \in \mathcal{M}$, and $\mathbf{n} \in \mathcal{M}$.
Chapter 12

Similarity

1. Matrix of a transformation. If $\mathcal{V}$ is a vector space with dimension $n$, then there is a one-to-one correspondence between the linear transformations from $\mathcal{V}$ into itself and the $n \times n$ matrices. The truth of this assertion is seen by ordering the elements of a basis of $\mathcal{V}$, like

$$\mathcal{B} = \{e_1, \ldots, e_n\},$$

and looking at the correspondence that pairs a transformation $T$ with the matrix $(\alpha_{ij})$, where

$$Te_j = \sum_{i=1}^{n} \alpha_{ij}e_i$$

for each $j \in \{1, \ldots, n\}$. This matrix $(\alpha_{ij})$ that corresponds to the transformation $T$ is referred to as the $\mathcal{B}$-matrix of $T$. Addition and multiplication of matrices is defined so that, when using this correspondence, products of matrices corresponds to composition of transformations, and sums of matrices corresponds to sums of transformations. In the special case when $\mathcal{V}$ is $\mathbb{R}^n$ or $\mathbb{C}^n$, and $\mathcal{B}$ is the standard basis, i.e. the basis whose $i^{th}$ vector contains 1 in the $i^{th}$ coordinate and 0 in all other coordinates, then the $\mathcal{B}$-matrix of $T$ is called the standard matrix of $T$.

A generalization of a $\mathcal{B}$-matrix allows us to obtain a correspondence between matrices and transformations like $T: \mathcal{V} \rightarrow \mathcal{W}$, where $\mathcal{V}$ and $\mathcal{W}$ are finite dimensional but not necessarily equal. If we are given such a linear $T$, a basis $\mathcal{B}$ of $\mathcal{V}$, a basis $\mathcal{C}$ of $\mathcal{W}$, and we order the two bases like

$$\mathcal{B} = \{e_1, \ldots, e_n\} \quad \mathcal{C} = \{f_1, \ldots, f_m\},$$

then we can associate $T$ with the matrix $(\alpha_{ij})$, where

$$Te_j = \sum_{i=1}^{m} \alpha_{ij}f_i$$

for each $j \in \{1, \ldots, n\}$. This matrix is called the $\mathcal{B}, \mathcal{C}$-matrix of $T$.

If $\mathcal{M}$ is a subspace of $\mathcal{V}$, and $\mathcal{N}$ is a complement of $\mathcal{M}$ in $\mathcal{V}$, then a transformation $T: \mathcal{V} \rightarrow \mathcal{V}$ gives rise to four other transformations that describe the action of $T$ in terms of its behaviour on the two subspaces $\mathcal{M}$ and $\mathcal{N}$. Starting with a vector $m \in \mathcal{M}$, the vector $T(m)$ can be written in terms of a basis $\mathcal{B} \cup \mathcal{C}$ of $\mathcal{V}$ as

$$T(m) = \alpha_1e_1 + \ldots + \alpha_ke_k + \beta_1f_1 + \ldots + \beta_if_i,$$

where $\mathcal{B} = \{e_1, \ldots, e_k\}$ is a basis of $\mathcal{M}$ and $\mathcal{C} = \{f_1, \ldots, f_l\}$ is a basis of $\mathcal{N}$. Two transformations, which encode the behaviour of $T$ on $\mathcal{M}$ are obtained by defining $T_{11}: \mathcal{M} \rightarrow \mathcal{M}$ and $T_{21}: \mathcal{M} \rightarrow \mathcal{N}$ by

$$T_{11}(m) = \alpha_1e_1 + \ldots + \alpha_ke_k \quad \text{and} \quad T_{21}(m) = \beta_1f_1 + \ldots + \beta_if_i.$$
Similarly, two other transformations \( T_{22} : \mathcal{N} \to \mathcal{N} \) and \( T_{12} : \mathcal{N} \to \mathcal{M} \) are defined, such that \( T_{22} \) takes a vector \( \mathbf{n} \in \mathcal{N} \) to the part of \( T(\mathbf{n}) \) that lives in \( \mathcal{N} \), and \( T_{12} \) takes \( \mathbf{n} \) to the part of \( T(\mathbf{n}) \) that lives in \( \mathcal{M} \). In this way, when a vector \( \mathbf{v} \in \mathcal{V} \) is written uniquely as \( \mathbf{v} = \mathbf{m} + \mathbf{n} \) with \( \mathbf{m} \in \mathcal{M} \) and \( \mathbf{n} \in \mathcal{N} \), one obtains the unique decomposition of \( T(\mathbf{v}) \) as a vector in \( \mathcal{M} \) plus a vector in \( \mathcal{N} \) via

\[
T(\mathbf{v}) = (T_{11}(\mathbf{m}) + T_{12}(\mathbf{n})) + (T_{21}(\mathbf{m}) + T_{22}(\mathbf{n})).
\]

This equation has a nice realization in terms of the \( \mathcal{B} \cup \mathcal{C} \)-matrix of \( T \), which can be divided into a \( 2 \times 2 \) block matrix

\[
A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix},
\]

where \( A_{11} \) is the \( \mathcal{B} \)-matrix of \( T_{11} \), \( A_{22} \) is the \( \mathcal{C} \)-matrix of \( T_{22} \), \( A_{21} \) is the \( \mathcal{B} \)-\( \mathcal{C} \)-matrix of \( T_{12} \), and \( A_{12} \) is the \( \mathcal{C} \)-\( \mathcal{B} \)-matrix of \( T_{12} \). Just as \( T(\mathbf{v}) \) is encoded in a matrix product, by dividing the \( \mathcal{B} \cup \mathcal{C} \)-coordinates of \( \mathbf{v} \) into two parts

\[
\mathbf{v} = (\epsilon_1 \mathbf{e}_1 + \ldots + \epsilon_k \mathbf{e}_k) + (\delta_1 \mathbf{f}_1 + \ldots + \delta_l \mathbf{f}_l),
\]

the formula above assumes the form

\[
\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \\ \delta_1 \\ \vdots \\ \delta_l \end{pmatrix} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_k \\ \delta_1 \\ \vdots \\ \delta_l \end{pmatrix} + \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} \delta_1 \\ \vdots \\ \delta_l \end{pmatrix}.
\]

The true power of this observation lies in the fact that the usual algebraic operations performed on scalar matrices work equally well on block matrices.

**Problem 1**

We continue with the notation introduced in the previous paragraphs. Assume that the transformations \( T : \mathcal{V} \to \mathcal{V} \) and \( U : \mathcal{V} \to \mathcal{V} \) have \( \mathcal{B} \cup \mathcal{C} \) matrices

\[
\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \quad \text{and} \quad \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix},
\]

respectively, where the \( 2 \times 2 \) block decompositions are given relative to the complement subspaces \( \mathcal{M} \) and \( \mathcal{N} \), as described above. Show that the \( 2 \times 2 \) decomposition of the \( \mathcal{B} \cup \mathcal{C} \)-matrix of \( T \circ U \) is

\[
\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix} \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} = \begin{pmatrix} A_{11} B_{11} + A_{12} B_{21} & A_{11} B_{12} + A_{12} B_{22} \\ A_{21} B_{11} + A_{22} B_{21} & A_{21} B_{12} + A_{22} B_{22} \end{pmatrix}.
\]

2. Coordinates and isomorphisms. An isomorphism \( \iota : \mathcal{V} \to \mathcal{W} \) is a linear transformation that is one-to-one and onto. If \( \mathcal{V} \) is a real (or complex) vector space of dimension \( n \), and \( \mathcal{B} = \{ \mathbf{e}_1, \ldots, \mathbf{e}_n \} \) is an ordered basis of \( \mathcal{V} \), then the basis determines an isomorphism \( \iota_{\mathcal{B}} \) with \( \mathbb{R}^n \) (or \( \mathbb{C}^n \)), by declaring that \( \iota_{\mathcal{B}} \) should pair the basis vectors in \( \mathcal{B} \) with the standard basis vectors of \( \mathbb{R}^n \) (or \( \mathbb{C}^n \)), then extending \( \iota_{\mathcal{B}} \) to force its linearity by defining

\[
\iota_{\mathcal{B}}(\sum_i \alpha_i \mathbf{e}_i) = \sum_i \alpha_i \iota_{\mathcal{B}}(\mathbf{e}_i).
\]

In this way, the vector \( \mathbf{v} = \alpha_1 \mathbf{e}_1 + \ldots + \alpha_n \mathbf{e}_n \) corresponds to the element

\[
\begin{pmatrix} \alpha_1 \\ \vdots \\ \alpha_n \end{pmatrix},
\]
and every question asked in the coordinate system determined by $B$ is converted to a question posed in standard coordinates.

**Problem 2** Assume $V$ is a vectors space of dimension $n$ and $B$ is an ordered basis of $V$. Show that the $B$-matrix of a transformation $T : V \rightarrow V$ has the same entries as the standard matrix of $\iota_B \circ T \circ \iota_B^{-1}$.

3. **Similarity.** As a basis $B$ is changed, most likely the $B$-matrix of the transformation $T$ will change with it, but since these potentially different matrices are representing the same transformation, we introduce a relation called **similarity** that lets us distinguish between, not individual matrices, but groups of matrices, so that different groups correspond to different transformations with genuine structural differences, not differences that arise because of the use of a different coordinate system. Consequently, two matrices are defined to be **similar** when one of the matrices coincides with the $B$-matrix of a transformation $T$, and the other matrix coincides with the $C$-matrix of the same transformation $T$, where $B$ and $C$ are bases of the domain of $T$. In terms of the isomorphisms $\iota_B$ and $\iota_C$, this is the same as saying one of the matrices is the standard matrix of $\iota_B \circ T \circ \iota_B^{-1}$, and the other is the standard matrix of $\iota_C \circ T \circ \iota_C^{-1}$.

**Problem 3** Show that the matrix $A$ is similar to the matrix $B$ if and only if there exists an invertible matrix $S$ so that $SA = BS$.

4. **Upper triangular form.** Mathematical induction is a common technique used to prove results in linear algebra because it is convenient and natural to induct on the dimension of the vector space. Every complex $n \times n$ matrix is similar to one in **upper triangular form**, i.e. it is similar to a matrix whose non-zero entries all lie on and above the main diagonal. The inductive proof of this assertion begins with the observation that, when $n = 1$, the statement is a triviality, and when $n > 1$, if one shows that the transformation associated with the matrix has an eigenvector, then we can conclude that the matrix is similar to one whose first column is all zeros, with the possible exception of the topmost entry. The induction hypothesis can then be used on the $(n - 1) \times (n - 1)$ southeast block. The proof of the triangular form assertion then rests on the existence of an **eigenvector**, i.e. a nonzero vector that is transformed to a multiple of itself.

The essential use of the complex scalar field occurs right now, when we attempt to show that every $n \times n$ matrix has an eigenvalue. Observe that $\alpha$ is an **eigenvalue** associated with a transformation $T$ exactly when $T - \alpha I$ is not invertible, and the nonzero vectors in the nullspace of $T - \alpha I$ are precisely the eigenvectors of $T$. One can then reason that an eigenvector exists if it is known that

$$(T - \alpha_1 I) \cdots (T - \alpha_k I) = 0$$

for some $\alpha_1, \ldots, \alpha_k \in \mathbb{C}$. The fundamental theorem of algebra implies the existence of such scalars $\alpha_i$ when it is known that $p(T) = 0$ for some polynomial $T$. Such a polynomial is found by considering that

$$\{e, T(e), \ldots, T^n(e)\}$$

must be a dependent set, and using this fact for several values of $e$.

**Problem 4** Assume $T : V \rightarrow V$ is a linear transformation of an $n$-dimensional complex vector space $V$. Show that there is a basis of $V$ relative to which the matrix of $T$ is upper triangular.

5. **Two variants on upper triangular form.** The immediate advantage of the upper triangular form is that the eigenvalues become visable on the main diagonal. One slight twist on the argument that gives the existence of a triangular form will yield a slightly stronger statement, which becomes a stepping stone in the proof of the Jordan canonical form.

**Problem 5** Assume $T : V \rightarrow V$ is a linear transformation of an $n$-dimensional complex vector space $V$, and $\alpha_1, \ldots, \alpha_k$ is an ordering of all the distinct eigenvalues of $T$. Show that there is a basis of $V$ relative
to which the matrix of $T$ is upper triangular, and the eigenvalues of $T$ appear on the diagonal in the order listed (so that $\alpha_i$ occurring before $\alpha_j$ on the diagonal implies $i \leq j$).

A second variant of this argument on the existence of a triangular form will show that two transformations can be simultaneously triangularized if it is known that the transformations commute. Two transformations $S$ and $T$ are **simultaneously triangularizable** when there exists a single basis $B$ relative to which the matrices of both $S$ and $T$ are upper triangular. If $S$ and $T$ commute, i.e. if $ST = TS$, then it is possible to show that $S$ and $T$ have a common eigenvector. This fact lets the same induction technique used in the previous two problems be utilized to prove the simultaneous triangularizability, and the result is a second stepping stone in the proof of a Jordan form.

**Problem 5** Assume $S, T : \mathcal{V} \rightarrow \mathcal{V}$ are both linear transformations of an $n$-dimensional complex vector space $\mathcal{V}$, and assume $ST = TS$. Show that there is a basis of $\mathcal{V}$ relative to which the matrices of both $S$ and $T$ are upper triangular.

**Problem 5** Let $\mathcal{P}$ denote the set of all rank one orthogonal projections in $\mathcal{B}(C^n)$. Show that the linear span of $\mathcal{P}$ is all of $\mathcal{B}(C^n)$.

**More Exercises**

1. Prove that the function that takes a transformation to its $\mathcal{B}$-matrix is one-to-one and onto.
2. Assume $\mathcal{L}$ denotes the set of linear transformations on an $n$-dimensional vector space and $M_n$ denotes the set of $n \times n$ matrices. Let $\Phi : \mathcal{L} \rightarrow M_n$ denote the function that takes a transformation $T$ to the matrix $\Phi(T)$, whose entries are the same as the entries of the $\mathcal{B}$-matrix of $T$. Show that $\Phi(ST) = \Phi(S)\Phi(T)$ for all $S, T \in \mathcal{L}$.
3. Assume $M$ denotes the matrix of a transformation $T$ relative to the standard basis

$$\{e_1, e_2, e_3, e_4\}$$

of $R^4$. Let $P$ denote the orthogonal projection onto the subspace $W$, which we define as the span of $C = \{e_1, e_2\}$. Let $B = \{e_3, e_4\}$ span the space $\mathcal{V}$. Show that the $2 \times 2$ matrix in the northeast corner of $M$ is the $\mathcal{B},\mathcal{C}$-matrix of the transformation $PT : \mathcal{V} \rightarrow W$.
4. Assume $M$ denotes the standard matrix of a transformation $T$ of $R^4$. Find a transformation $R$ and bases $\mathcal{B}$ and $\mathcal{C}$ so that the southwest corner of $M$ is the $\mathcal{B},\mathcal{C}$-matrix of $R$. 
Solutions
Chapter 13

Bases

13.1 Vector Spaces

Solution 1 If \( L \) is a subspace of \( V \), as defined above problem 1, then \( \alpha v + \beta w \in L \) whenever \( \alpha \) and \( \beta \) are scalars and \( v, w \in L \), so we need only prove the converse statement. With this goal, assume \( \alpha v + \beta w \in L \) for all scalars \( \alpha \) and \( \beta \), and all \( v, w \in L \). Since \( L \) is nonempty, there exists \( v \in L \), so \( 0v \in L \) by our assumption, and \( L \) contains the additive identity element. Also, it \( w \in L \), then \( (-1)w \in L \), so the additive inverse of every element in \( L \) also lies in \( L \). The rest of the mouthful that defines a vector space is now seen to hold for \( L \) because they hold for \( V \), and \( L \) is part of \( V \).

Solution 2 If \( v, w \in \text{span}(S) \) and \( \alpha, \beta \) are two scalars, then \( \alpha v + \beta w \in \text{span}(S) \), by the definition of \( \text{span}(S) \). We conclude that \( \text{span}(S) \) is a subspace that contains \( S \). Assume \( M \) is another subspace that contains \( S \), and assume \( v \in S \). Then \( v = \alpha_1 s_1 + \ldots + \alpha_k s_k \) with \( s_1, \ldots, s_k \in S \), and since \( S \) is inside \( M \), so are all of the vectors \( s_1, \ldots, s_k \), and \( M \) being a subspace implies that the linear combination \( v \) is also in \( M \), and \( \text{span}(S) \subseteq M \). Thus \( \text{span}(S) \) is the smallest subspace that contains \( S \).

Solution 3 We begin by proving the assertion in the paragraph preceding the problem, namely, if \( v = \sum_i \alpha_is_i \) is a non-zero vector in the span of \( S \) with \( \alpha_j \neq 0 \), and if \( T \) is the set obtained by adding \( v \) to \( S \) and removing \( s_j \) from \( S \), then \( S \) and \( T \) have the same span. A typical element in the span of \( T \) looks like

\[ \beta v + \sum_i \beta_i r_i, \]

with all \( r_i \in S \), and none of which equals \( s_j \). This is the same as the vector

\[ \sum_i \beta \alpha_i s_i + \sum_i \beta_i r_i, \]

which is clearly in the span of \( S \), so \( \text{span}(T) \subseteq \text{span}(S) \). Conversely, linear combinations of vectors in \( S \) that do not include the vector \( s_j \) must be in \( \text{span}(T) \), so assume we have a linear combination of the form

\[ \beta s_j + \sum_i \beta_i r_i, \]

with all \( r_i \in S \) and \( r_i \neq s_j \) for each \( i \). Solving the equation \( v = \sum_i \alpha_i s_i \) for \( s_j \), which is possible because \( \alpha_j \neq 0 \), and substituting into the equation above, gives the same vector in the form

\[ \frac{\beta}{\alpha_j} v - \sum_{i \neq j} \frac{\beta \alpha_i}{\alpha_j} s_i + \sum_i \beta_i r_i, \]
which is seen to be a linear combination of vectors in $T$, and hence is in $\text{span}(T)$. Thus we have $\text{span}(S) \subseteq \text{span}(T)$. This proves the assertion in the paragraph.

To prove the statement in problem 3, assume $S_1$ is an independent subset of $V$ with $n$ elements and $S_2$ is a spanning set in $V$. Write $S_1$ as

$$S_1 = \{s_1, \ldots, s_n\},$$

and since $S_2$ spans $V$, each vector in $S_1$ can be written as a combination of vectors from $S_2$. Begin by writing $s_1$ as a combination of vectors from $S_2$, such as

$$s_1 = \sum_i \alpha_i r_i,$$

with $r_i \in S_2$, and since $S_1$ is independent, we know that $s_1 \neq 0$ so $\alpha_j \neq 0$ for at least one of the indices $j$. Remove the vector $r_j$ from $S_2$ and add the vector $s_1$, resulting in a new set that we will call $S_3$, which has at most the same number of elements as $S_2$, and has the same span as $S_2$, so that $S_3$ spans the vector space $V$, since $S_2$ does.

The next step is to repeat the process with the spanning set $S_3$ and the independent set

$$\{s_2, \ldots, s_n\};$$

write $s_2$ as a combination of vectors from $S_3$, convicce ourselves that there must be a nonzero scalar coefficient next to a vector from $S_2$ in this combination (otherwise we would contradict the assumption that $S_1$ is independent). Remove the vector next to the nonzero scalar from the set, and add the vector $s_2$, resulting in a new set that we call $S_4$. This step is a special case of the general inductive step, which goes like this; assuming we have done this process $k$ times, we end up with a spanning set $S_{k+1}$, which has at most the same number of elements as $S_2$, and we have the independent set

$$\{s_k, \ldots, s_n\}.$$

Write $s_k$ as a combination of vectors from $S_{k+1}$,

$$s_k = \alpha_1 s_1 + \ldots + \alpha_{k-1} s_{k-1} + \sum_i \beta_i r_i,$$

where each $r_i \in S_2$. Notice that some $\beta_j$ must be nonzero, otherwise we would contradict the independence of $S_1$. Remove this corresponding vector $r_j$ from $S_{k+1}$ and add the vector $s_k$, giving us a new set $S_{k+2}$, which also spans $V$, and has at most the same number of elements as $S_2$. The independence of $S_2$ implies that this process will not stop before $n$ steps, which culminates in a set $S_{n+2}$, which has at most as many elements as $S_2$, and which contains all the vectors of $S_1$. It follows that $S_2$ has $n$ or more vectors.

**Solution 4**

Assume that $B_1$ and $B_2$ are two bases of $V$. Since $B_1$ spans $V$ and $B_2$ is independent, solution 3 shows that $B_1$ has at least as many elements as $B_2$. Interchanging the roles of $B_1$ and $B_2$ in this statement shows that $B_2$ has at least as many elements as $B_1$, so they must have the same number of elements.

**Solution 5**

Consider the subset of natural numbers

$$\{m \in N : \text{there is a spanning set of } V \text{ with } m \text{ elements}\}.$$

This set has a smallest element $n$, and there is a spanning set $B$ with $n$ elements. If any element of $B$ is removed, the resulting new set can no longer span $V$, since this new set will have fewer than $n$
elements. This means that each element of \( B \) fails to be a linear combination of the other elements of \( B \), which implies that \( B \) is independent.

**Solution 6** Consider the subset of natural numbers

\[
\{ m \in N : \text{there is an independent set containing } X \text{ with } m \text{ elements}\}.
\]

Every element of this set is less than or equal to the dimension of \( V \), by solution 3, so this set has a largest element, call it \( n \), and there is an independent set \( B \) containing \( X \) with \( n \) elements. We will show that \( B \) spans \( V \) by contradiction; if we assume that there is a vector \( v \) not in the span of \( B \), then the set \( B \cup \{ v \} \) is independent, and it contains \( X \). This would imply that \( n + 1 \) is in the above set of integers, contradicting that \( n \) is the largest element of the set. Thus \( B \) is a basis of \( V \) that contains \( X \).

**Solution 7** Assume \( N \) is a complement of \( M \) in \( V \), so by the definition in the paragraph preceding problem 7, there is a basis \( B = X \cup Y \) of \( V \), where \( X \cap Y = \emptyset \), \( X \) is a basis of \( M \), and \( Y \) is a basis of \( N \). If \( v \in M \cap N \), then \( v = \sum_{x \in X} \alpha_x x \) since \( v \) is in \( M \), and \( v = \sum_{y \in Y} \alpha_y y \) since \( v \) is in \( N \). It follows that

\[
\sum_{y \in Y} \alpha_y y = \sum_{x \in X} \alpha_x x,
\]

and

\[
\sum_{y \in Y} \alpha_y y - \sum_{x \in X} \alpha_x x = 0.
\]

This is a linear combination of the independent vectors in \( B \), so we must have all coefficients equal to zero. In particular,

\[
v = \sum_{x \in X} \alpha_x x = \sum_{x \in X} 0 \cdot x = 0,
\]

so \( M \cap N = \{ 0 \} \). Since \( B \) spans \( V \), so does \( X \cup Y \), and this union is a subset of \( M \cup N \), so \( M \cup N \) also spans \( V \).

Assume now that \( M \cup N \) spans \( V \) and \( M \cap N = \{ 0 \} \). Let \( X \) be any basis of \( M \), and let \( Y \) be any basis of \( N \). We define \( B \) to be the union of \( X \) and \( Y \), and we need to prove that \( B \) is a basis of \( V \). Assume we have a linear combination of vectors from \( B \) that comes out to be \( 0 \). This linear combination can be written

\[
\sum_{x \in X} \alpha_x x + \sum_{y \in Y} \alpha_y y = 0,
\]

which means

\[
w = \sum_{x \in X} \alpha_x x = \sum_{y \in Y} (-\alpha_y) y.
\]

Since \( w \) is in both \( M \) and \( N \), we must have \( w = 0 \) and

\[
o = \sum_{x \in X} \alpha_x x = \sum_{y \in Y} (-\alpha_y) y.
\]

Now the independence of \( X \) and \( Y \) imply that \( \alpha_x = 0 \) for all \( x \in X \) and \( \alpha_y = 0 \) for all \( y \in Y \), which establishes the independence of \( B \). To see that \( B \) spans \( V \), start with an arbitrary vector \( v \in V \), and use the assumption that \( M \cup N \) spans \( V \) to write \( v \) as a combination of vectors in \( M \cup N \). In this linear combination, write every vector from \( M \) as a combination of vectors in the basis \( X \), and write every vector from \( N \) as a combination of vectors in the basis \( Y \). What results is a linear combination of vectors from \( B \) that equals \( v \).

**Solution 8** Assume the hypothesis, and let \( B = X \cup Y \) be a basis of \( V \), with \( X \) a basis of the nullspace of \( T \). Let \( Y = \{ e_1, \ldots, e_k \} \), and let us prove that the set

\[
\{ T(e_1), \ldots, T(e_k) \}
\]
is independent. With this goal, assume that

$$\alpha_1 T(e_1) + \ldots + \alpha_k T(e_k) = 0,$$

so that

$$T(\alpha_1 e_1 + \ldots + \alpha_k e_k) = 0$$

and $$\alpha_1 e_1 + \ldots + \alpha_k e_k$$ is in the nullspace of $$T$$. But this vector is also in the span of $$Y$$, which is a complement of the nullspace, and problem 7 says that the only vector in both a subspace and it complement is the zero vector, which implies

$$\alpha_1 e_1 + \ldots + \alpha_k e_k = 0.$$  

By the independence of the vectors in $$Y$$, we conclude that $$\alpha_1 = \ldots = \alpha_k = 0.$$
Chapter 14

Similarity

Solution 9 not much here yet

Solution 10 not much here yet

Solution 11 not much here yet

Solution 12 Let $B = \{e_1, \ldots, e_n\}$ be a basis of $C^n$. For each $e_i \in B$, let $p_i$ be a polynomial for which $p_i(T)e_i = 0$. If $p = p_1 \cdots p_n$, then $p(T)e_i = 0$ for each $i = 1, \ldots, n$, so that the $B$-matrix of $p(T)$ is the zero matrix, and consequently $p(T) = 0$.

If $T : C^n \to C^n$ is linear, and $e$ is any vector in $C^n$, then problem 3 on page 118 implies that the vectors

$$e, T(e), T(T(e)), \ldots, T^{n-1}(e)$$

can not be independent, since there are no independent sets with more elements than a basis. Being dependent, there are scalars $a_0, a_1, \ldots, a_n$ for which

$$a_0e + a_1T(e) + \ldots + a_nT^n(e) = 0.$$

If $p$ is the polynomial $p(x) = a_0 + a_1x + \ldots + a_nx^n$, this equation can be written

$$p(T)e = 0,$$

which shows that given any vector $e$ there exists a polynomial $p$ with $e$ in the nullspace of $p(T)$.

Let $B = \{e_1, \ldots, e_n\}$ be a basis of $C^n$. For each $e_i \in B$, let $p_i$ be a polynomial for which $p_i(T)e_i = 0$. If $p = p_1 \cdots p_n$, then $p(T)e_i = 0$ for each $i = 1, \ldots, n$, so that the $B$-matrix of $p(T)$ is the zero matrix, and consequently $p(T) = 0$.

$$\begin{pmatrix} 1 & 0 \\ 0 & S \end{pmatrix} \begin{pmatrix} \alpha & A \\ 0 & B \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & S^{-1} \end{pmatrix}$$

Solution 13 waiting patiently for the proof to appear

Solution 14 Let $B = \{e_1, \ldots, e_n\}$ be a basis of $C^n$. For each $e_i \in B$, let $p_i$ be a polynomial for which $p_i(T)e_i = 0$. If $p = p_1 \cdots p_n$, then $p(T)e_i = 0$ for each $i = 1, \ldots, n$, so that the $B$-matrix of $p(T)$ is the zero matrix, and consequently $p(T) = 0$.

If $T : C^n \to C^n$ is linear, and $e$ is any vector in $C^n$, then problem 3 on page 118 implies that the vectors

$$e, T(e), T(T(e)), \ldots, T^{n-1}(e)$$

can not be independent, since there are no independent sets with more elements than a basis. Being
dependent, there are scalars \( a_0, a_1, \ldots, a_n \) for which

\[
a_0e + a_1T(e) + \ldots + a_nT^n(e) = 0.
\]

If \( p \) is the polynomial \( p(x) = a_0 + a_1x + \ldots + a_nx^n \), this equation can be written

\[
p(T)e = 0,
\]

which shows that given any vector \( e \) there exists a polynomial \( p \) with \( e \) in the nullspace of \( p(T) \).

Let \( \mathcal{B} = \{e_1, \ldots, e_n\} \) be a basis of \( C^n \). For each \( e_i \in \mathcal{B} \), let \( p_i \) be a polynomial for which

\[
p_i(T)e_i = 0. \quad \text{If } p = p_1 \cdots p_n, \text{ then } p(T)e_i = 0 \text{ for each } i = 1, \ldots, n, \text{ so that the } \mathcal{B}\text{-matrix of } p(T) \text{ is the zero matrix, and consequently } p(T) = 0.
\]

\[
\begin{pmatrix}
1 & 0 \\
0 & S
\end{pmatrix}
\begin{pmatrix}
\alpha & A \\
0 & B
\end{pmatrix}
\begin{pmatrix}
1 & 0 \\
0 & S^{-1}
\end{pmatrix}
\]

**Solution 15** Assume that the operator \( T \) is orthogonal to \( \mathcal{P} \). Writing \( T \) in its triangular form, we deduce that all
eigenvalues of \( T \), which appear on the diagonal of this triangular matrix, are zero, since \( \text{tr}(TP) = 0 \)
for each of the diagonal matrix units \( P \). It follows that \( \text{tr}(T) = 0 \), and also that \( \text{tr}(T^2) = 0 \).

The operator \( T^* + T \) is hermitian, and by the spectral theorem it is orthogonally diagonalizable, and hence in the span of \( \mathcal{P} \). It follows that

\[
\text{tr}(T(T^* + T)) = 0.
\]

Thus

\[
\text{tr}(TT^*) = \text{tr}(TT^*) + \text{tr}(T^2) = \text{tr}(T(T^* + T)) = 0.
\]

We then have \( 0 = ||TT^*|| = ||T||^2 = 0 \), so \( T = 0 \). Thus the span of \( \mathcal{P} \) is all of \( \mathcal{B}(C^n) \), since its perp
is \( \{0\} \).
Appendix
Appendix 1

Complex Numbers

1. **Endowing** \(R^2\) **with more structure.** The underlying set of complex numbers is the cartesian plane, which after endowing with a commutative multiplication, becomes the field \(C\) of complex numbers. An algebraic description of the multiplication is essential for computations and is given by the equation

\[
(\alpha_1, \beta_1) \ast (\alpha_2, \beta_2) \equiv (\alpha_1\alpha_2 - \beta_1\beta_2, \alpha_1\beta_2 + \beta_1\alpha_2).
\]

**Problem 1** Show that \((1, 0)\) is the multiplicative identity of this multiplication.

2. **The linearity of multiplication.** If \(v \in R^2\), then we can define a function

\[
T_v : R^2 \to R^2
\]

by \(T_v(u) = u \ast v\). In terms of the coordinates of \(u\) and \(v\) this definition assumes the form

\[
T_v(x, y) = \left( \begin{array}{c} \alpha \\ \beta \end{array} \right) \left( \begin{array}{c} x \\ y \end{array} \right),
\]

from which we see the linearity of each \(T_v\).

**Problem 2** Describe the geometric effect of the transformation \(T_v\) when

a. \(v = (0, 1)\)  b. \(v = (0, -1)\)  c. \(v = (1, 0)\)  d. \(v = (1, -1)\).

3. **The traditional notation.** After we are convinced that \((1, 0)\) is a multiplicative identity, we do not hesitate to employ the traditional symbolism that conveys this fact by writing \(1 \equiv (1, 0)\), and we will even go so far as to write \(j \equiv (0, 1)\) for any \(j \in R\). It is customary to denote the product of two complex numbers by juxtaposition: instead of writing \(u \ast v\) most folks write \(uv\). If we now write \(i \equiv (0, 1)\) we can uniquely express each complex number as \(\alpha + i\beta\), and since \(i^2 = -1\), we get

\[
(\alpha_1 + i\beta_1)(\alpha_2 + i\beta_2) = (\alpha_1\alpha_2 - \beta_1\beta_2) + i(\alpha_1\beta_2 + \beta_1\alpha_2),
\]

just as in high school.

**Problem 3** Plot the points \(1 + i\), \((1 + i)^2\), \((1 + i)^3\), and \((1 + i)^4\) in the plane, and give a description of what happens as \(n \to \infty\).

4. **Representations of fields.** A given field of scalars may have identical structural properties to a second scalar field; in such a situation we would be able to find an invertible function between these two fields that preserves all of their structure, and we would say the fields are **isomorphic**. For example, the field of real numbers is isomorphic to the field

\[
\{ \left( \begin{array}{cc} \alpha & 0 \\ 0 & \alpha \end{array} \right) : \alpha \in R \},
\]
and the function $\Psi$ that takes $\alpha$ to the matrix with $\alpha$ down the main diagonal is an invertible function that satisfies $\Psi(\alpha + \beta) = \Psi(\alpha) + \Psi(\beta)$ and $\Psi(\alpha\beta) = \Psi(\alpha)\Psi(\beta)$, for every $\alpha, \beta \in R$. In such a way we are able to consider various representations of the real numbers. The advantage of the diagonal matrix perspective of real numbers is that it allows us to see a larger number system that properly contains the real numbers, and the larger system is itself isomorphic to the complex number system. The larger system is

$$C \equiv \left\{ \begin{pmatrix} \alpha & \beta \\ -\beta & \alpha \end{pmatrix} : \alpha, \beta \in R \right\},$$

and the key to the following problem is to see which matrix is playing the role of $i$.

**Problem 4** Give an example of an invertible function $\Psi$ between $C$ and $C$ that satisfies

$$\Psi(u + v) = \Psi(u)\Psi(v) \text{ and } \Psi(uv) = \Psi(u)\Psi(v)$$

for all inputs $u$ and $v$.

1. 

5. Factoring complex polynomials. Another view of the complex numbers involves taking the real number field and appending to it a single new element, then generating the resulting enlarged scalar field from the requirement that sums and products should remain in the field. The simplest choice is to take $i$ to satisfy $i^2 = -1$, and append this number to the real number field. All possible sums and products can now be written in the form $\alpha + i\beta$, and suddenly the roots of all polynomials appear. The introduction of $i$ amounts to providing a root to the polynomial $x^2 + 1$, but the **fundamental theorem of algebra** tells us that every polynomial

$$a_nx^n + \ldots + a_1x + a_0$$

with complex coefficients can by written

$$a_n(x - r_1)\cdots(x - r_n),$$

with $r_1, \ldots, r_n$ denoting the roots of the polynomial.

**Problem 5** Use the quadratic formula to find the roots of the following polynomials, then factor the polynomial as $a_2(x - r_1)(x - r_2)$:

- a. $x^2 + x + 1$
- b. $x^2 + ix + 1$
- c. $ix^2 + x + 1$.

6. Operations on complex numbers. Of the various representations of complex numbers mentioned, the first one is most useful in giving a geometric flavor to the system. The addition and length of complex numbers behaves exactly as it does for vectors in $R^2$; thus if $\zeta = (\alpha, \beta) \in C$, then the **length** (also called the norm or absolute value) of $\zeta$ is obtained from the Pythagorean theorem $|\zeta| \equiv \sqrt{\alpha^2 + \beta^2}$.

2. Draw a picture in the plane that illustrates the use of the Pythagorean theorem as justification for the definition $|\zeta| \equiv \sqrt{\alpha^2 + \beta^2}$.

The projection onto the first coordinate is called the **real part**, and the projection onto the second coordinate is called the **imaginary part** of a complex number. In this way every complex number $\zeta$ is uniquely written as

$$\text{Re}(\zeta) + i \text{ Im}(\zeta),$$
where Re(\(\zeta\)) and Im(\(\zeta\)) denote the real and imaginary parts of \(\zeta\), respectively. Finally, there is the conjugation operation, which corresponds to a reflection across the horizontal axis: if \(\zeta = (\alpha, \beta) \in C\), then the conjugate of \(\zeta\) is \(\overline{\zeta} = (\alpha, -\beta)\).

**Problem 6** Let \(\zeta \in C\) denote an arbitrary complex number. Of the three scalars \(\zeta^2\), \(|\zeta|^2\), and \(\overline{\zeta}\), prove that two are equal and in general not equal to the third.

3. Assume \(p(x) = a_n x^n + \ldots + a_1 x + a_0\) with all coefficients \(a_i\) real numbers, and assume \(\zeta\) is a complex root; i.e. assume \(p(\zeta) = 0\). Prove that \(p(\overline{\zeta}) = 0\).

4. Give an example of complex numbers \(u\) and \(v\) such that \(u^2 + v^2 = 0\). Show that the corresponding matrix
   \[
   \begin{pmatrix}
   u & v \\
   -v & u
   \end{pmatrix}
   \]
   has no inverse. Finally, replace each \(u\) and \(v\) in this matrix with the \(2 \times 2\) matrices representing \(u\) and \(v\) (respectively) to obtain a \(4 \times 4\) matrix with real entries; determine the rank of this \(4 \times 4\) matrix.

5. Let \(u\) and \(v\) be complex numbers; show that every matrix of the form
   \[
   \begin{pmatrix}
   u & v \\
   -\overline{v} & \overline{u}
   \end{pmatrix}
   \]
   is invertible, with one exception.

6. Assume that
   \[
   h(\alpha, \beta, \gamma, \delta) = 
   \begin{pmatrix}
   \alpha & \beta & \gamma & \delta \\
   -\beta & \alpha & -\delta & \gamma \\
   -\gamma & \delta & \alpha & -\beta \\
   -\delta & -\gamma & \beta & \alpha
   \end{pmatrix},
   \]
   and let \(\mathcal{H} = \{h(\alpha, \beta, \gamma, \delta) : \alpha, \beta, \gamma, \delta \in R\}\); we call \(\mathcal{H}\) the set of **Hamiltonian Quaternions**. Give an example of a one-to-one function \(\Psi : C \rightarrow \mathcal{H}\) that satisfies
   \[
   \Psi(uv) = \Psi(u)\Psi(v) \text{ and } \Psi(u + v) = \Psi(u) + \Psi(v)
   \]
   for all \(u, v \in C\).

7. We adopt the notation of exercise 6. Show that there exist elements of \(\mathcal{H}\) that do not commute, but every non-zero element is invertible with its inverse in \(\mathcal{H}\).

7. Extending the domains of real functions.
Index
\(B\)-matrix, 26, 117

A generalization, 14
absolute value, 130
addition of vectors, 7
adjoint, 96, 102
An algebraic characterization of projections, 59
An algorithm, 41
axiom of choice, 114

Bases, 113, 121
basis, 8, 114
bilinear, 100
block diagonal form, 69
Block matrices, 68
block matrix, 60, 69
canonical form, 51, 86, 117
Canonical Forms, 85
Changing coordinates, 16
characteristic polynomial, 38
codomain, 1
outputs, 1
commutative diagram, 81
companion matrix, 90
complement, 115, 123
complete similarity invariant, 54
complex conjugate, 95
Complex Numbers, 129
complex polynomials, 71
complex vector space, 22
composition, 4
conjugate linear, 96
conjugate transpose, 96, 102
Coordinates, 13
coordinates, 15
\(B\)-coordinates, 15
counterexample, 3
cyclic transformation, 89
cyclic vector, 89
determinant, 36
diagonal matrix, 30
diagonalizable, 30
Diagonalization, 40
dilation, 64
Dilations, 64
dimension, 8, 114
domain, 1
inputs, 1
dot product, 55, 95
eigenspace, 47
Eigenspaces, 47
eigenvalue, 11
eigenvector, 11, 118
equivalent, 6

Fourier expansion, 96
free vector, 6
function, 1
  codomain, 1
domain, 1
equal, 1
input, 1
output, 1
range, 1

Functions, 1
fundamental theorem of algebra, 44, 86, 118, 130

Generalization, 70
Graham-Schmidt, 98
Graham-Schmidt orthogonalization, 99
guru
  urp, 131

Hamiltonian Quaternions, 131
Hermitian operators, 103

identity function, 4, 8
imaginary part, 130
implicit codomain, 4
implicit domain, 4
independent, 7
inner product, 95
Introduction, 1
invariant subspaces, 9

Invariants, 33
invariants, 34
invertible, 115
isometries, 106
isomorphic, 24, 53, 129
isomorphisms, 24
Isomorphisms and change of basis, 80

Jordan block, 91
Jordan blocks, 87
Jordan canonical form, 86, 117
Jordan form, 86

kernel, 32

length, 130
linear, 24
linear combination, 7, 113
linear transformation, 115, 117
linear transformations, 9, 24
linearly independent, 113
Lower triangular, 31

main diagonal, 30
mathematical induction, 99
Matrices, 23
matrices
    similar, 54
Matrix algebra, 28
multiplication transformation, 76
multiplicity, 45
nilpotent, 66
nontrivial subspace, 9
norm, 130
normal operator, 105
not one-to-one, 3
nullity, 32
nullspace, 32, 115
one-to-one, 3
onto, 2
ordered basis, 114
orthogonal reflection, 62
orthogonally diagonalizable, 104
orthonormal basis, 95
Partially defined functions, 4
partition, 12
Plane projections, 51
Plane reflections, 62
Polar Decomposition, 106
polar decomposition, 106
positive definite, 106
positive matrices, 106
positive operator, 106
positive operators, 106
positive part, 106
positive semidefinite, 106
Projections, 50
rank, 32, 115
Rank and nullity, 31
rank and nullity theorem, 32, 115
rational block, 91
rational canonical form, 86, 89–91, 117
Rational form, 89
Real and Imaginary Parts, 102
real and imaginary parts, 103
real part, 130
real polynomials, 71
real vector space, 22
reflection, 62
Reflections et cetera, 61
representations, 130
Rotations, 67
Running with the algebra, 19
scalar product, 95
scalar transformation, 45, 64
sequence, 2
Shears and nilpotents, 65
similar, 51, 117
similar transformations, 54
similarity, 51, 117
similarity class, 54
similarity invariants, 34
simultaneously diagonalizable, 31
simultaneously triangularizable, 31
simultaneously triangularized, 105
singular, 30
Some matrix forms, 30
Space projections, 55
Space reflections, 63
span, 7
spanning set, 113
spectral decomposition, 104
spectral theorem, 95
Spectral Theorem, 104
standard coordinates, 14
Standard inner product, 95
standard matrix, 26
subspace, 113
subspaces, 8

Tayloring coordinates to the function, 14
tensor product, 100
Tensors, 100
The fundamental theorem of algebra, 44
The leading role, 24
The matrix of a transformation, 26
trace, 34
transformations
  similar, 54
Transformations from one space to another, 79
transpose of a vector, 96
triangularizable, 31
Triangularization, 116, 124
trivial subspace, 9

unit disc, 10
unitary, 106
unitary equivalence, 95
Unitary Geometry, 94
unitary operator, 106
upper triangular, 31

vector, 6
vector field, 9
Vector fields, 9
Vector Space, 113
Vector Spaces, 113, 121
vectors
linear combination, 7
linearly dependent, 7
linearly independent, 7
span, 7
Vectors and Coordinates, 6
Visualizing transformations of the plane, 8
When polynomials are the vectors, 71
When the vectors are matrices, 75
When the vectors are solutions to D.E.’s, 78
zero vector, 6