Only Study Guide for MAT2611

LINEAR ALGEBRA

Department of Mathematical Sciences
University of South Africa

Contents

Preface

Welcome to the MAT2611 module. This module is a continuation of the MAT1503 module and deals with linear algebra.

Prescribed textbook

The prescribed textbook for this module is

Title:	**Elementary Linear Algebra** with Supplemental Applications
Edition:	10^{th} Edition, International Student Version
Authors:	Howard Anton and Chris Rorres
ISBN:	**978 – 0 – 470 – 56157 – 7**

It is absolutely essential that you purchase the prescribed textbook.

Linear algebra has a wide variety of applications, e.g. in the mathematical sciences, natural sciences and engineering. Linear algebra has become more popular in recent years because of the development of high speed computers, so that linear algebra is now often applied in numerical work.

Like all mathematics, linear algebra is best learnt by doing exercises **yourself.** This study guide consists of worked examples. **You must try these exercises yourself**, before you read our solutions. There are two kinds of exercises in the textbook from which we draw. The **Exercises**, which come first, are of a computational nature. The **Theoretical Exercises** require you to prove results.

NB: The numbering of chapters and sections in this study guide corresponds to the numbering in the textbook.

We hope you will enjoy the MAT2611 module and that this study guide will prove helpful.

Study Unit 1

Chapter 4: General vector spaces

The chapter is called "general vector spaces" to warn you that from now on, we will consider vector spaces other than R^n. The points in these vector spaces may not be points in the usual sense of R^n but could be things like functions!

1.1 Section 4.1: Real vector spaces

This definition is extremely important. If a set along with two operations satisfy these axioms then it is a vector space, by definition! We are used to the vector spaces R, R^2 and R^3 but many other sets and operations satisfy this definition. No matter, as long as the definition is satisfied, we have a vector space by definition.

Why "real" vector space? Because the allowed *scalars* in the definition are the real numbers, when we allow complex scalars we call it a *complex* vector space.

To illustrate the variety of vector spaces **Examples 1-6, 8** give some. **Example 2** is our old friend R^n while examples 2, 3, 4, 5, 6 & 8 give some new ones. Note that each such set with the given operations (which are of course no longer the usual plus and scalar multiplication of R^n) must also satisfy some other requirements such as having a zero vector (again, not the usual zero vector from R^n).

Note especially **Example 6** where each vector is an entire function!

Just to show that not everything that looks like it may satisfy the definition is a vector space, **Example 7** gives a set that is not a vector space. One of the axioms fails to hold. It can be quite time consuming to check if a given set with given operations forms a vector space or not as many of the exercises ask you to do.

Example 1 shows that we can make a vector space with one vector, the zero vector.

Theorem 4.1.1 gives some properties of the vector operations. These properties

were obvious for R^n but are not obvious in general. Remember all we may use to demonstrate these properties are the vector space axioms, hence the proofs look formal and awkward.

1.2 Section 4.2: Subspaces

We can have vector spaces as part of other vector spaces. When a part W of a vector space V itself satisfies the definition of vector space (under the same operations as V) we call it a subspace (of V). In principle, then, to check if a subset of V is a subspace of V we need to check the vector space axioms for this subset. Fortunately, most of these properties are inherited from V, for example, since elements of W are also elements of V they will satisfy 2, 3, 8, 9 of the definition. In fact, all we need to check to see if a subset W of V is a subspace of V is the following (**Theorem 4.2.1**)

- W must be non-empty (so you must be able to give at least one element in W)
- If you add any two vectors in W, you end up in W
- If you multiply any vector in W by any scalar, you end up in W.

The last two properties are called *closure* properties since when you do them you do not *leav*e the set, you stay inside, hence *closure*.

The concept of a subspace and this result is very important and will often be used.

Examples 1-10 give some subspaces and non-subspaces, in each case the properties above are checked. If they are all satisfied, we have a subspace, if not, we do not. Remember again that to show that any given property fails, any single counterexample will do. For example, in **Example 4** it is shown that $-1(\mathbf{v})$ for $\mathbf{v} = (1, 1)$ leaves the set. This is enough to show that the property fails. On the other hand, to show that the set is closed under the operations we may not use particular vectors but must use general vectors, such as in **Example 5**, where p and q are general vectors, their coefficients are not specified.

Theorem 4.2.4 shows that the set of solutions $\mathbf{x}$ to a system

$$A\mathbf{x} = \mathbf{0}$$

forms a subspace.

On p.183 *linear combination* is defined. This is a central concept in the rest of the course. A vector $\mathbf{w}$ is a linear combination of a set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ if you can write $\mathbf{w}$ as a sum of the vectors $\mathbf{v}_i$ with scalars in front, as in

$$\mathbf{w} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + ... + k_r\mathbf{v}_r.$$

Example 8 shows that any vector in R^3 is a linear combination of the vectors $(1,0,0),(0,1,0)$ and $(0,0,1)$. You are so used to this that you may not realize it. Make sure you understand this example.

To *check* if some vector $\mathbf{w}$ is a linear combination of a set of vectors $\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$ needs work. As **Example 9** shows you have to see if you can find coefficients $k_1,k_2,...,k_r$ to make $\mathbf{w}=k_1\mathbf{v}_1+k_2\mathbf{v}_2+...+k_r\mathbf{v}_r$ true. This always leads to a linear system in the k's. If you can solve it (that is, find k's) then it is a linear combination, if not, not. **Example 9** is very important. (The book often leaves out important steps for you to fill in and asks you to verify a step or statement). One cannot *see* that the system at the end of **Example 9** has no solution, set up an augmented matrix and check! You should get a contradictory last row, that is, one of the form

$$0\ 0\ |\neq 0$$

which reads "$0k_1+0k_2\neq 0$". This is clearly impossible, hence no solution.

Theorem 4.2.3 shows that the set of all linear combinations W of a set of vectors $\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$ in a vector space V, form a subspace of V. In other words, if you take all possible sums

$$\mathbf{w}=k_1\mathbf{v}_1+k_2\mathbf{v}_2+...+k_r\mathbf{v}_r$$

(all possible means that you let the k's range over all the real numbers) then you get a subspace of V. (choose appropriate k's to show that $\mathbf{0}$, each of the $\mathbf{v}_i$ and $\mathbf{v}_1+\mathbf{v}_2+...+\mathbf{v}_r$ are all in the subspace).

Also, according to **Theorem 4.2.3** W is the smallest subspace of V that contains $\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$. In other words, if you are a subspace and you contain $\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$ then you *have to* also contain all linear combinations of $\{\mathbf{v}_1,\ \mathbf{v}_2,\ ...,\mathbf{v}_r\}$. This is because of the closure requirement on subspaces - if you contain $\mathbf{v}_1$ and $\mathbf{v}_2$ you must be closed under $\mathbf{v}_1+\mathbf{v}_2$ hence you must contain $\mathbf{v}_1+\mathbf{v}_2$. You must also contain $k\mathbf{v}_1$ and $k\mathbf{v}_2$. This generalizes to force the subspace to contain all linear combinations of $\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r$.

The proof is important.

The set of all linear combinations of a set $S=\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$ is called the *space spanned* by S and denoted by span(S) or span$\{\mathbf{v}_1,\mathbf{v}_2,...,\mathbf{v}_r\}$. (We are justified in calling it a space as we have just noted that it is indeed a subspace.) In what sense does S span this space? S itself is just a puny finite set of vectors. However, if you imagine a linear combination of the vectors in S as something that you can make with just the vectors of S (and scalars), you can begin to see that, in some sense S does span this space - any vector in this space is reachable using only vectors from S (and scalars).

Example 12 looks at spaces spanned by one and two vectors in R^3.

Example 13 looks at a spanning set for P_n. What is a spanning set? It is a set that spans the space. So a spanning set for P_n will be a set of vectors S in P_n such that span(S) is the entire P_n. In some sense, these vectors are enough to "generate" the entire P_n.

The very important **Example 15** gives an example of 3 vectors that will not generate the entire R^3. To see whether the set of vectors S spans a space we must check whether *every* vector in the space can be written as a linear combination of the vectors in S. Hence, to see whether

$$(1,1,2),(1,0,1),(2,1,3)$$

span R^3 we must know if any vector (b_1,b_2,b_3) can be obtained as

$$k_1(1,1,2)+k_2(1,0,1)+k_3(2,1,3).$$

Hence we must always be able to solve for k_1,k_2,k_3 in

$$k_1(1,1,2)+k_2(1,0,1)+k_3(2,1,3)=(b_1,b_2,b_3).$$

Equating first, second and third components gives

$$\begin{aligned} k_1+k_2+2k_3 &= b_1 \\ k_1\qquad\quad +k_3 &= b_2 \\ 2k_1+k_2+3k_3 &= b_3 \end{aligned}$$

This leads to the augmented matrix:

$$\left[\begin{array}{cccc} 1 & 1 & 2: & b_1 \\ 1 & 0 & 1: & b_2 \\ 2 & 1 & 3: & b_3 \end{array}\right].$$

Gaussian elimination gives:

$$\left[\begin{array}{cccc} 1 & 1 & 2: & b_1 \\ 0 & -1 & -1: & b_2-b_1 \\ 0 & 0 & 0: & b_3-b_2-b_1 \end{array}\right].$$

The last row is a problem. It says that when

$$b_3-b_2-b_1\neq 0$$

we get a contradiction. Hence for such (b_1,b_2,b_3) there is no solution in terms of the k's and hence we cannot reach all of R^3 with these 3 vectors and hence they are not a spanning set for R^3 the way $(1,0,0),(0,1,0),(0,0,1)$ are. Note that many vectors $\mathbf{b}$ will not be in span $(1,1,2),(1,0,1),(2,1,3)$ e.g. $(1,1,1)$ - since here

$$\begin{aligned} & b_3-b_2-b_1 \\ = \;& 1-1-1 \\ = \;& -1 \\ \neq \;& 0. \end{aligned}$$

The book does not do Gaussian elimination but calls on a theorem and uses the determinant, which is also okay but less direct and less informative.

Theorem 4.2.5 says that two sets span the same space (that is - vectors we can reach as linear combinations of the one set are exactly those we can reach as linear combinations of the other) if and only if each of the vectors *in each set* is a linear combination of the vectors in the other set.

The proof is informative and not hard - try it.

Examples

(1) Let $V = \mathbb{R}^3$ under the operations of

$$\begin{aligned}(a,b,c)+(x,y,z) &= (a+x,\ b+y,\ c+z)\\ k(x,y,z) &= (kx,\ ky,\ kz).\end{aligned}$$

Then V is a vector space under these operations (check Definition 1).

Define

$$W = \left\{(a,b,2) \in \mathbb{R}^3 \mid a,b \in \mathbb{R}\right\}$$

under the same operations as in V. Then W is **NOT** a subspace of V.

Soln:

Observe that $0 \in V$ is of the form $(0,0,0)$. However $0 \neq 2$ and so $(0,0,0) \notin W$, i.e. $0 \notin W$ and so W is **NOT** a subspace of V.

Note: Observe that the elements of W are ordered triples $(x,y,z) \in \mathbb{R}^3$, where the third coordinate is equal to 2, i.e. $z = 2$. However in $(0,0,0) \in \mathbb{R}^3$, the third coordinate is $0 \neq 2$. That is why then $0 = (0,0,0) \notin W$. Hence W cannot be a subspace of V.

(2) P_2 is a subspace of P_3.

Soln:

$$P_2 = \left\{a_0 + a_1x + a_2x^2 \mid a_0, a_1, a_2 \in \mathbb{R}\right\}$$

and

$$P_3 = \left\{b_0 + b_1x + b_2x^2 + b_3x^3 \mid b_0, b_1, b_2, b_3 \in \mathbb{R}\right\}.$$

Then P_2 is a subset of P_3. Observe that

$$0 = 0 + 0x + 0x^2 \in P_2$$

i.e. the zero polynomial is in P_2. So $P_2 \neq \emptyset$. Also observe that

$$0 = 0 + 0x + 0x^2 + 0x^3 \in P_3$$

as well. Now let $u, v \in P_2$ and k be a scalar. Then

$$\begin{aligned}u &= a_0 + a_1x + a_2x^2,\\ v &= c_0 + c_1x + c_2x^2.\end{aligned}$$

So

$$\begin{aligned}u+v &= (a_0 + a_1x + a_2x^2) + (c_0 + c_1x + c_2x^2)\\ &= (a_0 + c_0) + (a_1 + c_1)\,x + (a_2 + c_2)\,x^2\\ &\in P_2\end{aligned}$$

$$\begin{aligned}ku &= k(a_0 + a_1x + a_2x^2)\\ &= ka_0 + (ka_1)x + (ka_2)x^2\\ &\in P_2.\end{aligned}$$

Thus the two conditions of the subspace theorem i.e. Theorem 6.2 are satisfied. Hence P_2 is a subspace of P_3.

(3) Let V be a vector space, W_1 and W_2 be subspaces of V. Then $W_1 \cap W_2$ is also a subspace of V.

Soln:

Since W_1 and W_2 are subspaces of V, they are nonempty subsets of V (by definition of a subspace). So $W_1 \cap W_2$ is also a subset of V. Now W_1 and W_2 are subspaces, which means that $0 \in W_1$ and $0 \in W_2$. So $0 \in W_1 \cap W_2$ and thus $W_1 \cap W_2 \neq \emptyset$. Let $x, y \in W_1 \cap W_2$ and k be a scalar. Then $x, y \in W_1$ and $x, y \in W_2$. Since W_1 and W_2 are subspaces, $x+y \in W_1$ and $x+y \in W_2$. So $x + y \in W_1 \cap W_2$. Also for k a scalar, $kx \in W_1$ and $kx \in W_2$. So that $kx \in W_1 \cap W_2$. Hence $W_1 \cap W_2$ is a subspace of V (by Theorem 4.2.1).

(4) Let $S = \{v_1, v_2, \ldots, v_k\}$ be a set of vectors in a vector space V and let W be a subspace of V which contains S. Then

(i) W contains span S.

Soln:

Since W contains S, we have that $v_1, v_2, \ldots, v_k \in W$. But W is a subspace of V and so (by the subspace theorem) W is closed under addition and scalar multiplication. Thus for all scalars $a_1, a_2, \ldots, a_k$, we obtain that

$$a_1 v_1, a_2 v_2, \ldots, a_k v_k \in W$$

and

$$a_1 v_1 + a_2 v_2 + \ldots + a_k v_k \in W.$$

Thus W contains all the linear combinations of

$$v_1, v_2, \ldots, v_k.$$

However span S is the set of all the linear combinations of $v_1, v_2, \ldots, v_k$. Hence W contains span S.

(ii) span S is a subspace of V.

Soln:

All elements in span S will be in V and so span S is a subset of V. Now observe that

$$0 = 0v_1 + 0v_2 + \ldots + 0v_k \in \text{span } S.$$

So span $S \neq \emptyset$. Now let $x, y \in$ span S. Then

$$\begin{aligned} x &= a_1 v_1 + a_2 v_2 + \ldots a_k v_k \\ y &= b_1 v_1 + b_2 v_2 + \ldots b_k v_k \end{aligned}$$

where $a_1, a_2, \ldots, a_k, b_1, b_2, \ldots, b_k$ are scalars. Then we obtain that

$$\begin{aligned} x+y &= (a_1v_1 + a_2v_2 + \ldots + a_kv_k) \\ &\quad + (b_1v_1 + b_2v_2 + \ldots + b_kv_k) \\ &= (a_1+b_1)\,v_1 + (a_2+b_2)\,v_2 + \ldots + (a_k+b_k)\,v_k \\ &\in \text{span } S \end{aligned}$$

i.e. $x+y$ is a linear combination of $v_1, v_2, \ldots, v_k$. Now for α a scalar, we obtain that

$$\begin{aligned} \alpha x &= \alpha\,(a_1v_1 + a_2v_2 + \ldots + a_kv_k) \\ &= \alpha a_1v_1 + \alpha a_2v_2 + \ldots + \alpha a_kv_k \\ &= (\alpha a_1)\,v_1 + (\alpha a_2)\,v_2 + \ldots + (\alpha a_k)\,v_k \\ &\in \text{span } S \end{aligned}$$

i.e. αx is a linear combination of $v_1, v_2, \ldots, v_k$. Hence span S is a subspace of V.

(iii) span S is the smallest subspace of V which contains

$$v_1, v_2, \ldots, v_k$$

i.e. any other subspace of V which contains $v_1, v_2, \ldots, v_k$ actually contains span S.

Soln:

Observe that

$$\begin{array}{ccccccccccccc} v_1 & = & 1\cdot v_1 & + & 0v_2 & + & 0v_3 & + & \ldots & + & 0v_{k-1} & + & 0v_k \\ v_2 & = & 0v_1 & + & 1\cdot v_2 & + & 0v_3 & + & \ldots & + & 0v_{k-1} & + & 0v_k \\ \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \vdots & & \\ v_k & = & 0v_1 & + & 0v_2 & + & 0v_3 & + & \ldots & + & 0v_{k-1} & + & 1\cdot v_k \end{array}$$

So each of $v_1, v_2, \ldots, v_k$ is a linear combination of

$$v_1, v_2, \ldots, v_k.$$

Thus $v_1, v_2, \ldots v_k \in$ span S. Call span $S = W$ and let W' be some other subspace of V which contains $v_1, v_2, \ldots, v_k$. Since W' is a subspace, it is closed under addition and scalar multiplication. Now since $v_1, v_2, \ldots, v_k \in W'$, W' contains all the linear combinations of v_1, v_2, $\ldots$, v_k. Thus W' contains all the elements of $W =$ span S. Hence W' contains span $S = W$.

1.3 Section 4.3: Linear independence

A non-empty set of vectors $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ in a vector space V is linearly independent if the only solution to

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + ... + k_r\mathbf{v}_r = \mathbf{0} \tag{1}$$

is the obvious one

$$k_1 = 0, k_2 = 0, ..., k_r = 0.$$

Note that making the k's 0 will make (1) true. The question is if there are any *other* solutions. If not, the set is linearly independent.

What we intend to mean with a set being linearly independent is that none of the vectors in the set is a linear combination of the other vectors in the set. Why then state it like (1), which seems awkward? Because (1) is an elegant way of saying that *none* is a linear combination of the others (**Theorem 4.3.1**). (If one was, say $\mathbf{v}_2 = k_1\mathbf{v}_1 + k_3\mathbf{v}_3 + ... + k_r\mathbf{v}_r$ then we need only bring $\mathbf{v}_2$ across the = sign to get

$$\mathbf{0} = k_1\mathbf{v}_1 - \mathbf{v}_2 + k_3\mathbf{v}_3 + ... + k_r\mathbf{v}_r$$

a non-trivial solution to (1))

So instead of saying that $\mathbf{v}_1$ cannot be written as a linear combination of the rest, and $\mathbf{v}_2$ cannot be written as a linear combination of the rest and $\mathbf{v}_3$ cannot.... we just say that the only solution to (1) is the zero solution, that covers all the cases.

Examples 1-3 illustrate the definition with two linearly dependent and one linearly independent set.

Examples 2-7 are important. They show how to test for linear independence: Basically we set up equation (1). This becomes a linear system which we write as an augmented matrix and try to solve. If the only solution is the zero solution the set is linearly independent. Let's do **Example 2** slowly:

Do the vectors

$$\mathbf{v}_1 = (1, -2, 3), \mathbf{v}_2 = (5, 6, -1), \mathbf{v}_3 = (3, 2, 1)$$

form a linearly dependent or independent set?

By the definition, we want to know if

$$k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + k_3\mathbf{v}_3 = \mathbf{0}$$

has any solutions except the zero solution.

Writing it out gives

$$k_1(1, -2, 3) + k_2(5, 6, -1) + k_3(3, 2, 1) = (0, 0, 0).$$

Equating components gives:

$$\begin{aligned} k_1 + 5k_2 + 3k_3 &= 0 \\ -2k_1 + 6k_2 + 2k_3 &= 0 \\ 3k_1 - k_2 + k_3 &= 0. \end{aligned}$$

As augmented matrix:

$$\left[\begin{array}{ccc:c} 1 & 5 & 3 & 0 \\ -2 & 6 & 2 & 0 \\ 3 & -1 & 1 & 0 \end{array}\right].$$

(Note that the vectors have become the columns of the matrix - useful to remember when you're in a hurry and want to skip steps- like in the exam!)

Gaussian elimination on $\left[\begin{array}{cccc} 1 & 5 & 3: & 0 \\ -2 & 6 & 2: & 0 \\ 3 & -1 & 1: & 0 \end{array}\right]$ gives:

$$\left[\begin{array}{cccc} 1 & 5 & 3: & 0 \\ 0 & 2 & 1: & 0 \\ 0 & 0 & 0: & 0 \end{array}\right].$$

We have a zero last row so we know there are non-trivial solutions and the vectors are linearly dependent. Indeed, setting

$$k_3 = t$$

we get

$$k_2 = -\frac{1}{2}t \text{ by the second row}$$

and

$$\begin{aligned} k_1 &= -5k_2 - 3k_3 \\ &= 2\frac{1}{2}t - 3t \\ &= -\frac{1}{2}t. \end{aligned}$$

For any $t \neq 0$ we will get a non-zero solution, e.g. $t = 2$ gives $-1, -1, 2$. Just to check, let's replace these k's in

$$k_1(1, -2, 3) + k_2(5, 6, -1) + k_3(3, 2, 1) = (0, 0, 0)$$

we get that indeed

$$-(1, -2, 3) - (5, 6, -1) + 2(3, 2, 1) = (0, 0, 0).$$

As an exercise, check whether

$$\mathbf{v}_1 = (1, -2, 3), \mathbf{v}_2 = (5, 6, 1), \mathbf{v}_3 = (3, 2, 1)$$

forms a linearly independent set. ($\mathbf{v}_2$ has been slightly changed).

Example 4 is different. When we set up the equation we get

$$\begin{aligned} a_0\mathbf{p}_0 + a_1\mathbf{p}_1 + a_2\mathbf{p}_2 + ...a_n\mathbf{p}_n &= \mathbf{0} \text{ or} \\ a_0x + a_1x^2 + ... + a_nx^n &= 0 \end{aligned}$$

where 0 is here the zero polynomial, that is, the polynomial $p(x) = 0$ for all x (the line lying on the x-axis in the xy-plane).

Now for there to be non-zero solutions to $a_0x + a_1x^2 + ... + a_nx^n = 0$ we would get some polynomial, for example

$$\frac{1}{2}x + 6x^2 + ... + 100x^n = 0 \text{ for all } x$$

This means this polynomial lies on the x-axis and is everywhere zero. This means in turn that every single value of $x \in R$ is a root! But it is well known that every polynomial of degree n with non-zero coefficients has at most n distinct roots, so it is not possible that every x value can be a root. Hence the set

$$\{x, x^2, ..., x^n\}$$

is linearly independent.

Theorem 4.3.1 shows what we discussed earlier: that a set is linearly independent if and only if none of the vectors is a linear combination of the others. The proof is important.

Examples 6 and **7** demonstrate the theorem by showing that we can write a vector as a linear combination of other vectors in those cases we found the sets to be linearly dependent.

Theorem 4.3.2 gives two small facts about linear independence and **Example 8** uses it to show that x and $\sin x$ are linearly independent.

Note that **Theorem 4.3.2** only holds for sets of 2 vectors. In general, linear dependence of a set $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_r\}$ does *not* mean that one of the vectors is a multiple of another. However, one must be a *sum* of multiples of the others.

Theorem 4.3.3 says that a set in R^n with more than n vectors must be linearly dependent. Roughly this is because we have, in the augmented matrix, fewer rows than columns and the augmented part is 0. We know from MAT1503 that such a system has infinitely many solutions, hence not only the zero solution.

We know almost nothing about checking sets of functions for linear independence - the topic of p.195-7. The discussion leads to **Theorem 4.3.4** which gives a criteria for a set of functions to form a linearly independent set. Roughly we take each function as the head of a column and put it's derivative under it, its second derivative under that and so on, $n-1$ times so we get a square matrix. We now work out the determinant of this matrix. It will generally be a function in x. If this function is not identically zero (that is, not everywhere zero), then the original functions are linearly independent.

1.4 Section 4.4 & 4.5: Basis and Dimension

We have intuitions about the "dimension" of different spaces, e.g. that a line has dimension 1 and that the world has dimension 3. What do we mean by this? Roughly, that 1 number is enough to give any particular point on a given line and that 3 numbers are enough to specify any particular point in space. So roughly, the dimension of a space is the number of coordinates we must give to specify an element (point) in that space. If we think carefully about this, these numbers we give actually specify the number of steps we need to take to reach the desired point. Once we realize this, we realize that behind the number is a size and a direction of the step we must take.

Example Giving the vector $(1,1)$ in R^2 only makes sense if we know this means "go one step to the right - that is, in the direction $(1,0)$ and then one step up -that is, in the direction $(0,1)$". So we actually are taking the vectors $(0,1)$ and $(1,0)$ for granted. The vectors $(1,0)$ and $(0,1)$ in terms of which any other vector can be given, are called the basis vectors.

For example, one step in direction $(0,1)$ in R^2 will not end up at the same point as one step in the direction $(1,0)$. Also, one step in the direction $(0,1)$ will not end up at the same point as one step in the direction $(0,2)$. We must therefore specify what the "steps" behind the coordinates are. These vectors (the vectors specifying what the unit steps are) will be a basis for the vector space. As **Figure 4.4.4** shows, we can use different "scales" or in effect bases, for a space. We could use (d) to specify a point in R^2 instead of (a). The coordinate vector $(1,1)$ in (d) would mean something different to $(1,1)$ in (a) though, since the "steps" (the basis vectors) differ.

Formally, a basis S for a vector space V is a *linearly independent set that spans* V.

Note then that V consists of all possible linear combinations of the vectors in S.

We have seen above that "S spans V" means that S is enough to specify (reach) any vector in V in the sense that any vector in V is a linear combination of those in S :

Example $S = \{(1,0,0),(0,1,0),(0,0,1)\}$ spans R^3 since any vector (a,b,c) in R^3 is a linear combination of those in S : since

$$(a,b,c) = a(1,0,0) + b(0,1,0) + c(0,0,1).$$

Given a basis, we therefore need only give the 3 component coordinate vector (a,b,c) to fully specify a point. For example

$$(-64, 1000, \sqrt{2}) = -64(1,0,0) + 1000(0,1,0) + \sqrt{2}(0,0,1).$$

On the other hand, no two vectors will do, that is no 2 vectors will span R^3.

Why do we require that the vectors are also linearly independent? **Theorem 4.4.1** gives one reason: If S spans V and is also linearly independent than we can write any vector in V in *only one way* as a linear combination of the vectors in S. This is very good as we do not want different representations for the same vector, this would cause unnecessary confusion. Say for example, we take S to span R^3 but not linearly independent, like $S = \{(1,0,0),(0,1,0),(0,0,1),(1,1,1)\}$. Now a vector will have 4 coordinates with respect to S and vectors will have more than one representation with respect to S. Indeed, for example

$$\begin{aligned}(1,1,1) &= 1(1,0,0) + 1(0,1,0) + 1(0,0,1) + 0(1,1,1) \text{ but also}\\ (1,1,1) &= 0(1,0,0) + 0(0,1,0) + 0(0,0,1) + 1(1,1,1).\end{aligned}$$

Hence $(1,1,1)$ has representations

$$(1,1,1,0) \text{ and}$$
$$(0,0,0,1).$$

This is not nice!

In some sense we do not *need* all 4 of the vectors in S, the first 3 are enough. So requiring linear independence ensures that we do not have a larger than necessary spanning set. Also, no smaller set will span V than the number in a basis, so a basis is a as small as possible spanning set.

We will often describe a vector space by just giving a basis. (we cannot give all the points in the space, there are usually infinitely many!) Also, although some spaces are easy to specify without giving a basis, such as R^3 (we all know what R^3 is) most are not, in which case it is easy to just give a basis as the definition of the space. So given only a basis S we know what V is, namely all possible linear combinations of vectors in S. So, in this sense, S specifies V.

Example 1 shows that the bases we have always used for R^n are in fact bases.

Example 3 shows that $S = \{(1,2,1),(2,9,0),(3,3,4)\}$ *also* forms a basis for R^n, so a vector space can have many bases! To show that S forms a basis we must show that S spans R^3 and is linearly independent. We know how to do this (see above). The book again calls on a theorem about determinants which is gives an alternative proof.

Example 9 shows how to find the coordinate vector of a vector w.r.t. a non-standard basis.

(a) That is, if we are given a vector $(5,-1,9)$ and our basis is

$$S = \{(1,2,1),(2,9,0),(3,3,4)\}$$

then the coordinate vector of $(5,-1,9)$ w.r.t S is the number of times we need to take the step $(1,2,1)$ then the step $(2,9,0)$ and then the step $(3,3,4)$ in order to reach $(5,-1,9)$. This means we want c_1, c_2, c_3 such that

$$c_1(1,2,1) + c_2(2,9,0) + c_3(3,3,4) = (5,-1,9).$$

This leads to a linear system that we solve with Gauss elimination, obtaining

$$(\mathbf{v})_S = (1,-1,2)$$

meaning that if we take the step $(1,2,1)$ once, then the step $(2,9,0)$ minus once (in other words, once in the opposite direction to $(2,9,0)$) then the step $(3,3,4)$ twice, we reach $(5,-1,9)$.

(b) If we are given the coordinate vector w.r.t to S and asked for the vector itself, it is easy. We are given the c_1, c_2, c_3 and just need to replace them in

$$c_1(1,2,1) + c_2(2,9,0) + c_3(3,3,4)$$

to get the vector.

Example 2 gives the standard basis for P_n - the set of polynomials of degree at most n, namely

$$\{1, x, x^2, ..., x^n\}.$$

Example 13 of section 4.2 showed that every vector in P_n (polynomial of degree at most n) can be written as a linear combination of these vectors. Further, **Example 4** of section 4.3 showed them linearly independent, hence by definition, they do form a basis for P_n. Finding coordinates polynomials w.r.t. this basis is very easy, just read of the coefficients of the x powers as (b) shows.

Example 4 gives the standard basis for $n \times n$ matrices, you should be happy with this basis too.

Example 3: Say you are given a linearly independent set S and are asked for a basis for span S. The set S is given as linearly independent and by definition, *spans* span S! So S itself is a basis.

We will give a method of finding a basis later in case S is *not* linearly independent.

Page 204 defines finite-dimensional and infinite dimensional vector spaces and **Example** 6 gives some examples. We have mostly worked with finite dimensional spaces up to now. (Note that we have not yet defined dimension properly!)

Theorem 4.5.2 states formally what we discussed briefly - that

(a) If we take more vectors than are in a basis, we get linear dependence

(b) If we take fewer vectors than are in a basis, we can no longer reach all the vectors in the space.

Theorem 4.5.1 at last gives the result that whatever basis we choose, it will have the same number of vectors. Now we can define dimension of a vector space V as the number of vectors in (any) basis for V.
Example 3 gives an example of finding a basis and the dimension of the solution space to a homogeneous system. We get two spanning vectors and since they are linearly independent (for example, the second components of $\mathbf{v}_1$ and $\mathbf{v}_2$ are 1 and 0 respectively, no scalar multiple of $\mathbf{v}_2$ will give $\mathbf{v}_1$) they form a basis for the set of solutions

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{bmatrix}$$

of the system.

Theorem 4.5.3

(a) If S is a linearly independent set and you add a vector $\mathbf{v}$ not in span S, then $S \cup \{\mathbf{v}\}$ is still linearly independent.

(b) If you delete from a (linearly dependent) set S a vector $\mathbf{v}$ that is a linear combination of the other vectors in S, then the remaining vectors span the same space without $\mathbf{v}$ as with $\mathbf{v}$. That is, you do not need $\mathbf{v}$ to reach all of span S, the other vectors are enough.

Theorem 4.5.4 says that for n vectors in a space of dimension n we need not check both conditions for being a basis. That is, if they are linearly independent then they will also span the space and if they span the space they will also be linearly independent. Either one of the conditions ensures the other.

Example 6(a) uses this **Theorem** to show that $\{(-3,7),(5,5)\}$ form a basis for R^2 by noting their linear independence.

(b) first notes that

$$(2,0,-1) \text{ and } (4,0,7)$$

are linearly independent (since the only multiple of the first that will give the second is 2 but then the third coordinate is incorrect) - hence they are not scalar multiples of each other. Since both have y component 0, no vector with a non-zero y component will be a linear combination of these two. Hence if we add $(-1,1,4)$ we still get a linearly independent set by the the Plus/Minus theorem. Then by **Theorem 4.5.4** we need not check that they span R^3 since R^3 has dimension 3 and we have 3 linearly independent vectors.

Theorem 4.5.5

(a) states that we can always obtain a basis from any spanning set by removing vectors, if necessary, from the spanning set.

(b) states that we can always obtain a basis from a linearly independent set by, if necessary, adding vectors to the set.

Theorem 4.5.6 states that the dimension of a subspace is at most the dimension of the space itself. Also, if the dimensions are the same, then the subspace must in fact be the entire space.

1.5 Section 4.6: Change of basis.

Since different bases may be suitable for different purposes we often use different bases for the same vectors space. Recall that for vector $\mathbf{v}$ in a vector space V and a basis

$$S = \{\mathbf{v}_1, \mathbf{v}_2, ...\mathbf{v}_n\}$$

the coordinate vector of $\mathbf{v}$ w.r.t. S is those k's such that

$$\mathbf{v} = k_1\mathbf{v}_1 + k_2\mathbf{v}_2 + ... + k_n\mathbf{v}_n.$$

The coordinate vector is then

$$(\mathbf{v})_S = (k_1, k_2, ..., k_n).$$

We will often write this as a column vector:

$$\begin{bmatrix} k_1 \\ k_2 \\ . \\ . \\ k_n \end{bmatrix}.$$

We would then generally work with the coordinate vector instead of the vector itself.

When we change bases, it stands to reason that the coordinate vector will change, how?

Let two bases for a two dimensional space be

$$\begin{aligned} S &= \{\mathbf{u}_1, \mathbf{u}_2\} \\ T &= \{\mathbf{v}_1, \mathbf{v}_2\}. \end{aligned}$$

Say we have the coordinate of a vector $\mathbf{w}$ w.r.t S, hence we have

$$[\mathbf{w}]_S = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}$$

such that

$$\mathbf{w} = a_1\mathbf{u}_1 + a_2\mathbf{u}_2.$$

What is the coordinate vector of $\mathbf{w}$ w.r.t. T?

First write $\mathbf{u}_1, \mathbf{u}_2$ w.r.t. T, say

$$\begin{aligned} \mathbf{u}_1 &= c_1\mathbf{v}_1 + c_2\mathbf{v}_2 \\ \mathbf{u}_2 &= d_1\mathbf{v}_1 + d_2\mathbf{v}_2 \end{aligned}$$

Then

$$\begin{aligned} \mathbf{w} &= a_1\mathbf{u}_1 + a_2\mathbf{u}_2 \\ &= a_1(c_1\mathbf{v}_1 + c_2\mathbf{v}_2) + a_2(d_1\mathbf{v}_1 + d_2\mathbf{v}_2) \\ &= (a_1c_1 + a_2d_1)\mathbf{v}_1 + (a_2c_2 + a_2d_2)\mathbf{v}_2 \end{aligned}$$

Hence

$$[\mathbf{w}]_T = \begin{bmatrix} a_1c_1 + a_2d_1 \\ a_2c_2 + a_2d_2 \end{bmatrix}$$

which can be written as:

$$\begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}.$$

Which is very nice, since this means that, whatever $[\mathbf{w}]_S$ is, we need only multiply it on the left with

$$P = \begin{bmatrix} c_1 & d_1 \\ c_2 & d_2 \end{bmatrix}$$

to get $[\mathbf{w}]_T$. So the matrix efficiently changes one coordinate vector to another. We now note that the column $\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}$ is $[\mathbf{u}_1]_T$ (since $\mathbf{u}_1 = c_1\mathbf{v}_1 + c_2\mathbf{v}_2$) and $\begin{bmatrix} d_1 \\ d_2 \end{bmatrix}$ is $[\mathbf{u}_2]_T$ (since $\mathbf{u}_2 = d_1\mathbf{v}_1 + d_2\mathbf{v}_2$). Hence, in order to find P we need only write each of the basis elements of S in terms of the basis T. The coordinate vector of the first basis element found is then column 1 of P, that of the second column 2 of P and so on.

This holds in general for basis with n-elements.

Example 1 gives an example of this result and **Example 2** reverses the directions, that is, we want the matrix that changes the coordinate from the T basis to the S basis.

Theorem 4.6.2 gives the relation between these 2 examples. If P is the transition matrix from S to T then P^{-1} is the transition matrix from T to S.

1.6 Section 4.7: Row space, Column space and Nullspace.

Given a matrix, there are three spaces we can associate with it:

The row-space: This is simply the span of the row vectors, in other words, all linear combinations of the rows.

The column-space: This is simply the span of the column vectors, in other words, all linear combinations of the columns.

The nullspace is a little different, it is the set of solutions x to the system

$$Ax = 0.$$

Theorem 4.7.1 states that a vector b is in the column space of a matrix A if and only if the system $Ax = b$ is consistent (that is, has a solution). This is clear once you realize the very important fact that the product Ax is simply x_1 times the first column of A plus x_2 times the second column of A and so on. This is a very important fact!

Example 2 uses the existence of a solution to show that a vector is in the column space, clearly this very solution will give the vector as a linear combination of the column vectors.

Theorem 4.7.2 says the following: Given a system $Ax = b$, if I have *any* solution x_0 to it, than any *other* solution can be written as x_0 plus a linear combination of the basis vectors for the *nullspace* of A. Conversely, any vector that can be written as x_0 plus a linear combination of the nullspace basis vectors *is* a solution to the system $Ax = b$.

Example 3 illustrates this theorem by first showing that the general solution to the system $Ax = b$ can be written as the sum of any particular solution, in this case

$$\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ \frac{1}{3} \end{bmatrix}$$

to $Ax = b$ plus a linear combination of the nullspace basis vectors

$$\begin{bmatrix} -3 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -4 \\ 0 \\ -2 \\ 1 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} -2 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \end{bmatrix}.$$

How do we find bases for the row space, column space and nullspace of a matrix?

Nullspace: Solve the homogeneous system. The vectors without the parameters form a basis.

Theorem 4.7.4 states that row-operations do not change the row space of a matrix. In other words span of the original rows is the same as span of rows in row echelon form.

The discussion following the proof notes that row operation *can* change the *column* space. However, the discussion goes on to show, row operations do not change linear relations between columns. For example, if

$$c_1 = 2c_2 + 4c_3$$

before we do row operations, then, after row operations when the columns have become c_1', c_2', c_3' we will still have

$$c_1' = 2c_2' + 4c_3'$$

This is important and will be used.

Theorem 4.7.6 states that if A and B are row-equivalent matrices, then the column vectors of A are linearly independent if and only if those of B are. Also, a particular set of columns of A form a basis for the column space of A if and only if the corresponding columns of B form a basis for the columns space of B.

Theorem 4.7.5 Since, in row-echelon form, the rows are linearly independent (and obviously span the row space!), it must be the case that the rows in row-echelon form, form a basis for the row space. (Can you see why the rows in row-echelon form are linearly independent?)

Also, for a matrix in row-echelon form, the columns containing leading ones of rows, form a basis for the column space.

Example 5 gives a matrix in row-echelon form and uses this theorem to give bases for the row space and column space. It is easy to read of the non-zero rows. It is a little more effort to see which columns contain leading ones, but in this case they are column 1 (contains row 1's leading one) column 2 (contains row 2's leading one) and column 4 (contains row 3's leading one).

Example 6 & 7 are very important and gives the general method of finding bases for row and column spaces.

Row space: Since row operations do not change the row space and the rows in row-echelon form, form a basis for the row space, we do gauss elimination on the matrix up to row-echelon form. The non-zero rows then form a basis for the row space.

For the column space: We can also use row-echelon form. Since row operations do not change linear dependencies amongst the columns, a linear independent set of columns in row-echelon form will have a corresponding linear independent set of columns in the original matrix. So we take these original columns. (clearly the linearly independent columns form a basis as they obviously span the column space!)

Example 8 asks for a basis for span $\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3, \mathbf{v}_4\}$. This is the same as asking for a basis for the row space of a matrix with these vectors as rows. Hence we can proceed as above.

Example 6 gives an example where we want a basis for the row space of a matrix but we insist that this basis consist of original rows, that is, we want a subset of the original rows which form a basis. Note that when we found a basis for the column space in **Example 7** we got original columns for our column space basis, so we need only transpose the matrix and find a basis for the column space of the transposed matrix, exactly as in **Example** 6.

Example 10 gives 5 vectors and

(a) wants a *subset* of them which forms a basis for their span. This is exactly like example 7 except the vectors are not given as rows of a matrix. So we do the same as example 7, make these vectors the *columns* of a matrix and find a basis for the column space.

(b) We must express the vectors not in the basis as linear combinations of those in the basis. Because row operations do not change linear relationships amongst columns, we can look at the columns in *reduced* row-echelon form (*reduced* row echelon form because it makes is particularly easy to see linear relations amongst the columns in this form) . Here

$$\{\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_4\}$$

form a basis and we must express

$$\mathbf{w}_3, \mathbf{w}_5$$

as linear combinations of them.

It is easy to see that

$$\mathbf{w}_3 = 2\mathbf{w}_1 - 1\mathbf{w}_2$$

hence also

$$\mathbf{v}_3 = 2\mathbf{v}_1 - 1\mathbf{v}_2$$

(which would have been much harder to see using the $\mathbf{v}$'s directly!)

Also, it is easy to see that

$$\mathbf{w}_5 = \mathbf{w}_1 + \mathbf{w}_2 + \mathbf{w}_4$$

hence also

$$\mathbf{v}_5 = \mathbf{v}_1 + \mathbf{w}_2 + \mathbf{w}_4.$$

1.7 Section 4.8: Rank and nullity

There are 4 vector spaces of interest for any matrix A

- Row space of A
- column space of A
- nullspace of A
- nullspace of A^T.

On the face of it, the row space and column space have very little to do with each other, it is therefore quite surprising that:

Theorem 4.8.1 They have the same dimension. This is because

1) we have seen that the *dimensions* of the row and column spaces of a matrix A are unchanged under row operations.

2) In row-echelon form, the number of non-zero rows equals the number of leading ones (clearly, every non-zero row implies a leading one, and every leading one implies a non-zero row, hence the numbers are equal). Since the number of rows gave the row-space dimension and the number of leading ones gave the column-space dimension, these two must be equal.

This common number is called the *rank* of A. The dimension of the nullspace (that is - solve the homogeneous system and see how many parameters (or basis vectors) you get) is called the *nullity* of A.

Example 1 finds the rank and nullity of a matrix A. To find the rank we can calculate the row space or column space. Row space is easy, just do row operations until (reduced) row echelon form and count the non-zero rows.

For the nullspace, note that the row operations will give the same row-echelon form, except that you carry a column of 0s in the last column. Since there are only two leading variables all the others are parameters, hence we get 4 parameters and 4 vectors in the solution to the homogeneous system. Hence the nullity is 4.

Theorem 4.8.7 remarks that the rank of A equals the rank of A^T.

It was not a coincidence in **Example 1** that rank plus nullity equaled the number of columns, whence **Theorem 4.8.3** which states that this is always the case. Remember: n = number of *columns* (not rows)

Theorem 4.8.3 explains why the

$$rank(A) + nullity(A) = n$$

if A has n columns and **Theorem 4.8.4** relates the rank and nullity to the form of the solution to $Ax = 0$, namely that

(a) $rank(A) = number$ of leading variables in the solution

(b) $nullity(A) = number$ of parameters in the general solution.

The matrix in **Example 1** has 6 columns hence $rank + nullity = 6$.

Example 3 gives the rank and the number of columns and finds the nullity.

Clearly, since the columns of an $m \times n$ matrix are in R^m and the rows in R^n and the dimensions of R^m and R^n are m and n respectively, the dimension of the column space cannot be more than m and that of the row space cannot be more than n. Since these two are equal, the rank cannot be more than $\min(m, n)$.

Example 4 illustrates this point for a 5×7 matrix.

Theorem 4.8.4 has a name which means the writer considers it important. We have the following equivalences:

$$\begin{array}{rcl} Ax & = & b \text{ is consistent (that is, has a solution)} \\ & \leftrightarrow & \\ & & b \text{ is in the column space of } A \\ & \leftrightarrow & \\ & & \text{the rank of } A \text{ does not increase when you add the extra column } b. \end{array}$$

The idea is that, since the rank is the number of linearly independent columns, if b is in the column space, it is not linearly independent of the original columns, and hence the rank stays the same. On the other hand, if the rank stays the same after adding b, then the number of linearly independent columns has not gone up. Hence b must be linearly dependent on the original columns.

Theorem 4.8.4 also gives three equivalences for an $m \times n$ matrix A:

$$\begin{array}{rcl} Ax & = & b \text{ is consistent for } \textit{every } b \text{ (that is, has a solution)} \\ & \leftrightarrow & \\ & & \text{the column space of } A \text{ is the entire } R^m \\ & \leftrightarrow & \\ & & \text{the rank of } A \text{ equals } m. \end{array}$$

The proof is not difficult and is similar to the previous proof.

An overdetermined system $Ax = b$ is one with more equations than unknowns, this means there are more rows than columns. The b column is thus bigger than

the number of columns of A and hence, since there are not enough columns to span R^m (the space in which b lives) by **Theorem 4.8.4** the system cannot be consistent for each b.

Theorem 4.8.5 says that if $Ax = b$ is a consistent system (with a solution) of m equations in n unknowns and A has rank r, then the general solution consists of $n - r$ parameters (remember all solutions consist of a particular solution plus the general solution to the homogeneous system $Ax = 0$).

Example 6 gives an example of this.

Theorem 4.8.4 gives some equivalent conditions on a homogeneous system having only the trivial solution:

$$\begin{aligned} Ax &= 0 \text{ has only the trivial (zero) solution} \\ &\leftrightarrow \\ &\text{the column vectors of } A \text{ are linearly independent} \\ &\leftrightarrow \\ &Ax = b \text{ has at most one solution for every } b. \end{aligned}$$

A linear system with more unknowns than equations is called an under determined system. An under determined system need not have any solutions. For example, the system

$$\begin{aligned} x + y + z &= 1 \\ x + y + z &= 2 \end{aligned}$$

has no solution, even though there are 3 unknowns and 2 equations. However, if an under determined system does have a solution, it must have many, since, by **Theorem 4.8.5** there must be at least one parameter.

Example 6 gives an example of this.

Theorem 4.8.10 summarizes much of the work so far. Make sure you understand it.

Study Unit 2

Chapter 5: Eigenvalues, Eigenvectors

If we multiply a vector x with a matrix A, as in Ax, we know that what we get is a linear combination of the columns of A specified by x (x_1 times the first column plus x_2 times the second column ...plus x_n times the n-th column).

Clearly, in general the answer we get will in no way resemble x! In those peculiar cases where Ax merely causes x to be multiplied by some scalar λ, that is

$$Ax = \lambda x$$

we call λ an *eigenvalue* of A and x the corresponding *eigenvector*.

Geometrically, for $x \in R^n$, x is merely stretched ($\lambda > 1$) or shrunk ($1 > \lambda > 0$) or turned around and stretched ($\lambda > -1$) or shrunk ($-1 < \lambda < 0$).

Such eigenvectors/eigenvalues have many important properties and applications, many of which you will come across in your math courses.

2.1 Section 5.1: Eigenvalues and eigenvectors

We start off with the definition written out nicely, with fig.5.1.1 showing what various λ do geometrically. Note that we exclude $\lambda = 0$ from the eigenvalues, since trivially, we always will have that $A0 = \lambda 0$. This says nothing about the matrix A so we ignore this case. (Note that this does not exclude $\lambda = 0$ from being an allowed eigenvalue for some *non-zero* x, since $Ax = 0x$ for non-zero x is an interesting, non-trivial case.)

Example 1 verifies that a given vector/scalar is an eigenvector/eigenvalue pair by showing that, for this λ and x, we indeed have that $Ax = \lambda x$.

We note that for a given eigenvalue, it is always the case (not only above) that any non-zero scalar multiple of an eigenvector is again an eigenvector. This is not hard to see: If

$$\begin{aligned} Ax &= \lambda x \text{ then} \\ A(cx) &= c(Ax) = c(\lambda x) = \lambda(cx). \end{aligned}$$

Since

$$A(cx) = \lambda(cx)$$

cx is also an eigenvector. However, of course, the eigen*values* are unique. (If any scalar multiple of a $\lambda \neq 0$ was also an eigenvalue then all λ would be eigenvalues!)

The book first illustrates the finding of eigenvalues only. A polynomial of degree n has at most n roots, and the characteristic equation of an $n \times n$ matrix will be an n-th degree polynomial so an $n \times n$ matrix has at most n eigenvalues.

Example 3 finds the eigenvalues for a 3×3 matrix. Looking at the characteristic equation it is certainly not obvious what the roots are! If we believe that the author was nice enough to set up A to have at least one integer solution (remember polynomials can have any type of solutions, rational fractions, irrational, complex!) then we can start by looking at the constant term c_n, that is,the term without λ's. In this case the integer solution must be a divisor of c_n - because c_n must be the product of all the terms x_i in the factors $(\lambda - x_i)$. We can then start by listing all factors of c_n and testing to see whether they are solutions. In this case 4 works so we can now divide by $(\lambda - 4)$ to get a quadratic equation, for which we know how to find the roots.

Example 4 shows that the eigenvalues of an upper-triangular matrix are just the diagonal entries. This follows from Theorem 2.1.2 which says that the determinant of a triangular matrix is just the product of the diagonal entries.

Theorem 5.1.2. is the general statement of this fact for $n \times n$ triangular matrices, that their eigenvalues are just the diagonal entries. So we will be very happy if the given matrix is diagonal!

Example 5 illustrates this for a lower triangular matrix. The discussion that follows reassures us that we will not be asked to find eigenvectors in case they are complex, although complex eigenvalues and eigenvectors are entirely possible for matrices with real entries.

Theorem 5.1.3 is a summary of the results of the chapter so far and is important.

2.1.1 Finding Eigenvectors and Bases for Eigenspaces

We have not yet discussed how to find eigenvectors. Since eigenvectors have been defined as the non-zero vectors in the *null space* of

$$\lambda I - A$$

for given A and we know that the null-space of a matrix is a vector space, we will call this null space the *eigenspace*. Note that all vectors in the eigenspace are then eigenvectors, except for the zero vector.

So since any non-zero linear combination of the vectors in the nullspace basis of $\lambda I - A$ will be an eigenvector, it is this basis we give when asked for eigenvectors.

Example 5 is the first time we find both eigenvalues and eigenvectors. The steps are

· Form the matrix $\lambda I - A$
· Work out the determinant of $\lambda I - A$ (that is, find the characteristic equation)
· Find the roots of this determinant (these are the eigenvalues)
· Replace each root λ_i into $\lambda I - A$ and find a basis for its nullspace
∗∗(these are the eigenspaces associated with each respective eigenvalue)

The operations in each of the steps are familiar - except you have seldom worked out a determinant in a variable and may need some practice in polynomial long division when finding roots!

This example is extremely important, as you will be doing this often!

Powers of a matrix: The discussion gives the interesting fact that once we know the eigenvalues of a matrix A, we know them for any power of A - they are just the corresponding power of the eigenvalue. Furthermore the eigenvectors stay the same!

Theorem 5.1.4 states this explicitly. Example 8 is a direct application of this theorem.

Theorem 5.1.5 states that a square matrix A is invertible if and only if 0 is not an eigenvalue of A. The proof is not hard.

Example 9 illustrates the use of this theorem.

Theorem 5.1.6 is our newest set of statements equivalent to the invertibility of A. This list grows throughout the book.

2.2 Section 5.2: Diagonalization

In this section we will be interested in finding bases for R^n which consist of the eigenvectors of a given matrix A. As the book states, such special bases tell us about A and makes certain calculations involving A easier. Such bases also have special significance in many contexts.

The matrix diagonalization problem: We have the very important equivalence between

The eigenvector problem:

Given $n \times n$ matrix A, does there exist a basis for R^n consisting of eigenvectors of A?

The matrix diagonalization problem:

Given $n \times n$ matrix A, does there exist an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix?

If the second condition is met, we call A *diagonalizable* and say that P diagonalizes A.

Theorem 5.2.1 proves that these conditions are equivalent. The proof is important and not tricky - make sure you understand it.

How does one diagonalize an $n \times n$ matrix A? As follows:

Step 1: Find the eigenvectors of A and hope that you get n linearly independent ones. (If not, A is simply not diagonalizable)

Step 2: Take these very eigenvectors and make them the columns of a matrix, call it P.

You are now assured that

$$P^{-1}AP$$

will be a diagonal matrix, with A's eigenvalues on the diagonal, in the same order as you placed the eigenvectors (that is, if you put λ's eigenvector in column i then $P^{-1}AP$ will have λ in column i.) Note that, unless asked, you need *not* work out $P^{-1}AP$ since you are assured that it will equal the diagonal matrix with eigenvalues on the diagonal.

How do we know whether A will have n linearly independent eigenvectors or not? The straightforward way is to find the eigenvectors and see whether they are n linearly independent vectors.

Example 1 diagonalizes the given matrix A. That is, we find for the given A a matrix P that diagonalizes it. We find the bases for the eigenspaces. For $\lambda = 2$ we get as basis for the nullspace of $\lambda I - A$ or here $2I - A$:

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

and for $\lambda = 1$ we get as basis for the nullspace $\lambda I - A$ or here $I - A$:

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

If we assume that these three are linearly independent (which will only follow from Theorem 5.2.2, later) then we can form the matrix P by simply making these the columns of P, so

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

We are now assured that

$$P^{-1}AP = D = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

that is the first two columns contain the eigenvalue 2 since the first two columns of P consisted of the eigenspace basis for 2 and the last contains a 1 since the last column of P consisted of the eigenspace basis for 1.

(Comment: Once you have found P you need not check that $P^{-1}AP = D$, the diagonal matrix with the eigenvalues on the diagonal)

You could have put the columns in P in any order - D would then change accordingly.

Are all matrices diagonalizable? No. Example 2 gives one that is not. When we find bases for the eigenspaces of this matrix we only get 2 vectors in total. Since we need 3 to form P the matrix cannot be diagonalized. In the alternative solution we replace the eigenvalue 1 in $\lambda I - A$ and after row operations, note that the rank of the matrix $\lambda I - A$ is 2 and hence the nullity is only 1. This means the basis for the nullspace (the eigenspace of 1) contains only one vector. The same holds for the eigenspace corresponding to 2. So we can deduce, without doing the work of finding the eigenvectors, that we are going to get too few vectors (2 instead of 3) and can conclude that the matrix is not diagonalizable. (Note that we would need to get the matrices $\lambda I - A$ into row echelon form in order to see how many non-zero rows (rank) we get.)

Theorem 5.2.2 says that the eigenvectors of different eigenvalues, are linearly independent.

The remark that follows generalizes this to when eigenvalues have one or more eigenspace basis vectors . For example, if λ_1 has eigenspace basis

$$\{u_1, u_2\}$$

and λ_2 has eigenspace basis

$$\{v_1, v_2, v_3\}$$

and λ_3 has eigenspace basis

$$\{w_1, w_2\}$$

then the entire set

$$\{u_1, u_2, v_1, v_2, v_3, w_1, w_2\}$$

will still be linearly independent. This generalization is not proved.

Theorem 5.2.3: Since Theorem 5.2.2 assures us that eigenvectors of different eigenvalues, are linearly independent, we can conclude that if an $n \times n$ matrix A has n *different* eigenvalues we will get a set of n linearly independent eigenvectors and then by Theorem 5.2.1 A will be diagonalizable.

If therefore we find n different eigenvalues we can conclude that A is diagonalizable without the work of finding the diagonalizing matrix P. This is what Example 3 and 4 do.

Note that Theorem 5.2.3 does not say that if we do *not* have n different eigenvalues that A is *not* diagonalizable, only that if we have n different ones, it is.

We have touched on the case when we do not get n different eigenvalues. Example 6 looks at two matrices. Firstly, the 3×3 identity matrix. Since we are interested in solutions to

$$Ax = \lambda x$$

so here

$$Ix = 1x$$

which holds for all vectors in R^3! So we can certainly choose 3 independent vectors. (Of course, I is already as diagonal as can be so it is a bit artificial to pretend to check whether its diagonalizable. (For any diagonal matrix D, there exists a P for which

$$P^{-1}DP$$

is diagonal, namely $P = I$!))

Secondly the matrix

$$J_3 = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

also has the repeated eigenvalue $\lambda_{1,2,3} = 1$. But now

$$J_3x = 1x$$

does not have all vectors as solutions. When we look at the matrix

$$1I - J_3$$

we get

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

which is in row-echelon form, so we can see the rank is 2 and nullity thus 1. Hence there is only one basis vector in the eigenspace of 1 and we need 3 linearly independent eigenvectors, we conclude that J_3 is not diagonalizable..

(The textbook calculates the nullspace and finds only one vector in its basis which is also okay, but a little more work)

After this example the book discusses when exactly a matrix is diagonalizable. The algebraic multiplicity of an eigenvalue is the number of times it appears as a factor in the characteristic equation, so, for example, the algebraic multiplicity of 3 in $(\lambda - 3)^2(\lambda - 1)$ is two while that of 1 is one. The geometric multiplicity is the number of vectors in the basis for the eigenspace of the given eigenvalue. This cannot be deduced from the characteristic equation. You have to plug the eigenvalue in the matrix equation $\lambda I - A$ and find the nullspace. By Theorem 5.2.5,

(a) the geometric multiplicity cannot be more than the algebraic multiplicity and

(b) if the geometric multiplicity is equal to the algebraic multiplicity for each eigenvalue, then the matrix is diagonalizable, and conversely.

Hence, if you find, for an eigenvalue of algebraic multiplicity 3 say, only 2 independent vectors in the solution of the associated nullspace, then you can conclude that the matrix is *not* diagonalizable.

The next paragraph shows how easy it is to find powers of a matrix once it is diagonalized, indeed, if

$$\begin{aligned} P^{-1}AP &= D \text{ then,} \\ (P^{-1}AP)^k &= D^k \\ &= P^{-1}A^kP \text{ (as noted in the text), hence} \\ A^k &= PD^kP^{-1} \end{aligned}$$

and D^k is just the matrix D with the diagonal entries to the k-th power.

Example 5 uses this method to find A^{13} for a diagonalizable matrix A.

Study Unit 3

Chapter 6

We will not do all of Chapter 6.

3.1 Section 6.1: Inner products

We are used to the *dot product* between two vectors, namely

$$\begin{aligned} & (a_1, a_2, ...a_k) \cdot (b_1, b_2, ..., b_k) \\ = \quad & a_1b_1 + a_2b_2 + ... + a_kb_k. \end{aligned}$$

Much like we could extract the crucial features from R^n and define an abstract vector space, we can extract the crucial features from the dot product for R^n to define what we mean by a dot product on other vector spaces, such as function spaces. The definition gives us this: A binary operation on a vector space (whatever the vectors are) that satisfies the definition is called an *inner product.*

We will generally only use the dot product and leave you to read the rest of section 1. If we need an inner product for vector spaces other than R^n we will introduce it as needed.

Note that the inner product is used to *define* length and distance in a vector space, that is, we need an inner product to define the length of a vector or the distance between two vectors. Exactly how are size and distance defined?

length of vector $\mathbf{u}$, denoted $\|\mathbf{u}\|$, is defined as

$$\|\mathbf{u}\| \quad = \quad \langle \mathbf{u}, \mathbf{u} \rangle^{\frac{1}{2}} \text{ (that is, the root of } \mathbf{u}\text{'s dot product with itself).}$$

distance between vectors $\mathbf{u}$ and $\mathbf{v}$, denote by $d(\mathbf{u}, \mathbf{v})$, is defined as

$$d(\mathbf{u}, \mathbf{v}) \quad = \quad \|\mathbf{u} - \mathbf{v}\| \text{ (that is, the length of their difference).}$$

Note that with this definition we are in a position to calculate lengths and distances in cases were we have no intuitive idea of what either *mean*, for example, what on earth would "the distance between two polynomials" mean? Or two continuous functions? With this definition we can calculate them! See **Example 7**.

Example 6 defines an inner product for 2×2 matrices and **Example** 8 defines an inner product for the space of polynomials P_2. **Example 9** defines an inner

product for $C[a,b]$ the space of continuous functions on the interval $[a,b]$. Interestingly, the inner product is defined as the integral of the product of the two functions.

3.2 Section 6.2: Angle and orthogonality in inner product spaces

It is easy to imagine what we mean by the *angle* between two vectors when we are working in R^2 or R^3 but what about other vector spaces, such as spaces of polynomials or continuous functions? It is impossible to visualize what we would mean by the angle between two polynomials! In section 6.2 we again define angle abstractly by extracting the properties of the usual angles in R^n. This leads to the definition of angle on p.346 where angle is defined in terms of the relation between the dot product of two vectors and the product of their norms (sizes)! To be exact, this gives the cos of the angle, so we must still find arccos of the answer obtained.This gives a definition we can use whenever we have defined a dot product on a vector space, even though we cannot visualize it. Of course, for this to be a generalization of the concept of angle in R^n this more abstract definition must give our familiar angles when we apply it to R^n and indeed it does, as remarked as the bottom of p.309 for R^2 and R^3.

In R^2 and R^3 two vectors were orthogonal if they pointed in perpendicular directions, and the test was whether their dot product was 0. Hence we can define an abstract concept of orthogonal the same way. As long as we have an inner product (the generalized dot product) in the space we can use this abstract definition. We can then define two vectors $\mathbf{u}, \mathbf{v}$ to be orthogonal if

$$\langle \mathbf{u}, \mathbf{v} \rangle = 0.$$

How do orthogonal matrices or polynomials or functions look then? **Examples 3 and 4** give examples of two vectors which turns out to be orthogonal for 2×2 matrices and polynomials in P_2. By their definition of inner product

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

are orthogonal in M_{22} and

$$x \text{ and } x^2$$

are orthogonal in P_2. Note that it does not make sense to try to see that they actually point in perpendicular directions!

Since the Pythagoras' **Theorem** is defined in terms of distances, and we can define distances in terms of dot products, we can give a generalized Pythagoras' **Theorem** for arbitrary vector spaces which have inner products.

Example 5 illustrates the validity of this general Pythagorean **Theorem** for the polynomials x and x^2.

We know what it means for a vector $\mathbf{v}$ to be orthogonal to a single vector $\mathbf{u}$ namely $\langle \mathbf{u}, \mathbf{v} \rangle = 0$. If we have a subspace W of an inner product space V, then

we say $\mathbf{v}$ is orthogonal to the *subspace* W if $\mathbf{v}$ is orthogonal to *every vector in* W. The set of all vectors that are orthogonal to W is called the *orthogonal complement* of W and denoted $W^{\perp}$.

Theorem 6.2.4 gives some properties of the orthogonal complement of a subspace W

It is itself a subspace

The only vector that W and $W^{\perp}$ have in common is 0.

The orthogonal complement of $W^{\perp}$ is again W.

Theorem 6.2.5 shows that, interestingly, the row space and nullspace of a matrix are orthogonal complements of each other under the usual dot product. (remember that there may be other valid inner products definable on vectors in R^n under which this need not be the case.) Also, the column space of A and the nullspace of A^T are orthogonal complements of each other.

Since the orthogonal complement of a subspace is again a subspace, we can find a basis for it. **Example** 6 finds a basis for the orthogonal complement of the space spanned by the vectors $\mathbf{w}_1, \mathbf{w}_2, \mathbf{w}_3, \mathbf{w}_4$.

To do this we can use Theorem 6.2.5. That is, we make these vectors the rows of a matrix (then their span is exactly the matrix's row space) so we need only find the nullspace!

3.3 Section 6.3: Orthonormal bases, Gram-Schmidt Process

(we will not discuss QR-decomposition - and it will not be examined)

We have defined what it means for two vectors to be orthogonal. A *set* of vectors is called orthogonal if all the pairs of the vectors in the set are orthogonal. If, further, each of the vectors has norm 1, the set is called *orthonormal*.

Example 1 gives an example of an orthogonal set in R^3. Each pair of the vectors has inner-product zero with each other.

To transform any vector into one with the same direction but having norm 1, we need merely divide the vector by its norm, as is explained after example **1**.

Example 2 normalizes the vectors of example **1** in order to obtain an ortho*normal* set.

We introduce the concept of orthonormal because orthonormal *bases* are very nice to work with. So working with an orthonormal basis will save us a lot of effort compared to an arbitrary basis.

Theorem 6.3.2 explains why. If we are given a vector $\mathbf{u}$ and we want the coordinates of $\mathbf{u}$ w.r.t. an orthonormal basis, we need merely take the inner

product of **u** and each of the basis vectors!

Example 3 gives an application of **Theorem 6.3.1**.

Theorem 6.3.2 shows that if we work with an orthonormal basis, then we can find norms, distances and inner products the way we would for R^n by working with the coordinate vectors. This means that even if we are in a strange vector space with strange definitions of norm, distance and inner product, if we work with an orthonormal basis, we can use the familiar norm, distance and inner product on the coordinate vectors of the actual vectors.

Example 4 illustrates that we really can work with the coordinate vectors to get the norm instead of getting the norm directly (which may be much more effort).

We can use Theorem 6.1.1 and Theorem 6.3.2 together to find coordinates w.r.t. orthogonal bases. Firstly, we normalize the basis vectors. Then use Theorem 6.3.2 which says we need only take inner products with each of the basis vectors, then use Theorem 6.1.1 to take the norm outside the bracket. This leaves the vector as a linear combination of the original basis before normalizing.

Theorem 6.3.1 says that orthogonal vectors are linearly independent.

Example 5 gives 3 orthonormal vectors, which since they are linearly independent by the Theorem above, also span R^3 by being 3 linearly independent vectors in 3 dimensional space. Hence they form a basis for R^3.

Theorem 6.3.4 states that every vector **u** in a vector space V can be written as an element of any finite dimensional subspace W plus an element of $W^{\perp}$. Also, there is only one way to do this. We call the components the *orthogonal projection of* **u** *on* W and the *component of* **u** *orthogonal to* W respectively.

Theorem 6.3.4 gives formulas for calculating orthogonal projections, all that is needed is the inner product.

Example 6 gives one such calculation.

Theorem 6.3.5 says that we can always find orthonormal bases for finite-dimensional vector spaces.

The steps to find such a basis are given in the proof. They change a given basis to an orthonormal one. We will use this process extensively later so make sure you can follow the recipe, an example of which is **Example 7**.

We will not discuss QR-decomposition.

3.4 Section 6.4: Best approximation; least squares.

Not part of the work to be examined.

Study Unit 4

Chapter 7: Diagonalization

4.1 Section 7.1: Orthogonal matrices

As you know it is hard work to find the inverse of a matrix. In the very special case that we need merely transpose a matrix to get its inverse, we call it an orthogonal matrix. That is, A is orthogonal if $A^{-1} = A^T$.

Example 1 gives one such matrix by showing that if we multiply the matrix with its transpose, we get the identity matrix.

Example 2 shows that rotation matrices are orthogonal.

Theorem 7.1.1 gives conditions for a matrix to be orthogonal. A is orthogonal if and only if the row vectors form an orthonormal set if and only if the column vectors form an orthonormal set.

Theorem 7.1.2 gives some properties of orthogonal matrices:

The inverse of an orthogonal matrix is also orthogonal.

A product of orthogonal matrices is orthogonal.

If A is orthogonal then its determinant can only be 1 or -1.

We will not discuss Orthogonal matrices as linear operators.

Sometimes we want to change between two orthonormal bases.

Theorem 7.1.5 says that in this case, the transition matrix will be orthogonal.

Examples 4 and 5 give applications of the theorem.

4.2 Section 7.2: Orthogonal diagonalization

This section starts with two problems: (Note the similarity with those at the start of Section 7.1)

The orthonormal eigenvalue problem:

Given an $n \times n$ matrix A, can we find an orthonormal basis for R^n consisting of eigenvectors of A?

The orthogonal diagonalization problem:

Given an $n \times n$ matrix A, is there an *orthogonal* matrix P such that $P^{-1}AP$ is diagonal?

(In other words, is A not only diagonalizable, but diagonalizable by a special orthogonal matrix P? Up to now we place no conditions on the diagonalizing matrix P.) Recall that an orthogonal matrix P is one of which the columns are orthonormal.

The second problem immediately raises the question of which matrices are so diagonalizable and how to find such orthogonal P. Firstly it is noted that only symmetric matrices A could possibly be so diagonalized. Indeed, if P is orthogonal, then $P^{-1} = P^T$ hence

$$P^{-1}AP = P^TAP = D \text{ by assumption}$$

so

$$A = PDP^T$$

since D is diagonal, also $D = D^T$ so

$$\begin{aligned} A^T &= (PDP^T)^T \\ &= (P^T)^T D^T P^T \\ &= PDP^T \\ &= A \end{aligned}$$

hence A must be symmetric. So A must be symmetric to be so diagonalized. Conversely, can we orthogonally diagonalize *any* symmetric A? Theorem 7.2.1 answers in the affirmative by showing that A is orthogonally diagonalizable $\leftrightarrow$ A has an orthonormal set of n vectors$\leftrightarrow$ A is symmetric.

Okay, now how do we find orthogonally diagonalizing P for symmetric matrices?

First we need Theorem 7.2.2 which says that the eigenvalues of a symmetric matrix are real (not complex) and that eigenvectors from different eigenspaces are orthogonal.

So we can now give the procedure for orthogonally diagonalizing a symmetric matrix:

- Find a basis for each eigenspace of A (set up the characteristic equation,

solve for the eigenvalues, replace back in $\lambda I - A$ and get bases for the nullspace)

- Apply the Gram-Schmidt process (if necessary!) to each of the bases to obtain an orthonormal basis for each eigenspace.
- Form the matrix P which consists of the basis vectors built in the previous step. This matrix will orthogonally diagonalize A.

Remarks: If an eigenspace has only one vector in the basis, you need only normalize the vector, you do not have to do the Gram-Schmidt process.

Example 1 finds a P that orthogonally diagonalizes the given A. This is an important example as you must be able to orthogonally diagonalize matrices.

Study Unit 5

Chapter 8: Linear Transformations

Linear transformations are the natural maps between vector spaces in that they preserve the crucial properties of vector spaces. More particularly, they have as image something like their domain. What exactly this means will become clearer as we proceed.

Before you start reading this part of the guide, please read

Chapter 4.9 – 4.11

as background. These sections discuss linear transformations for the important special case where the domain and codomain are the familiar R^n and R^m. In these cases linear transformations have nice geometrical representations. Also many of the results there will carry over to the more general case where our vector spaces need not be R^n.

So, just to make it clear, we will be working with linear transformations between arbitrary vector spaces V and W, that is, transformations

$$T : V \to W$$

where V and W can be any vector spaces.

What is a linear transformation? Let us first consider the case

$$T : R^n \to R^m.$$

That is T takes vectors $\mathbf{x}$ in R^n to vectors $\mathbf{w}$ in R^m. If each component of the vector $\mathbf{w} = (w_1, w_2, ..., w_m)$ is some linear combination of the components of $\mathbf{x} = (x_1, x_2, ..., x_n)$, in other words

$$\begin{aligned} w_1 &= a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n \\ w_2 &= a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n \\ &\vdots \\ w_m &= a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n \end{aligned}$$

then we say that T is a linear transformation.

That is, $T : R^n \to R^m, T(\mathbf{x}) = \mathbf{w}$ is a linear transformation if and only if if each of the "new" components w_i of $\mathbf{w}$ is a linear combination of the old components x_1 to x_n of $\mathbf{x}$.

Note that we use a's with subscripts because we do not have enough letters(!), and also because this notation already looks like a matrix product - indeed we can then write:

$$\begin{aligned} \mathbf{w} &= A\mathbf{x} \text{ where} \\ A &= \begin{bmatrix} a_{11} & a_{12} & .. & .. & a_{1n} \\ a_{21} & a_{22} & .. & .. & a_{2n} \\ . & . & . & . & . \\ . & . & . & . & . \\ a_{m1} & a_{m2} & . & . & a_{mn} \end{bmatrix}. \end{aligned}$$

This shows that

T is a linear transformation from R^n to R^m if and only if $T(\mathbf{x}) = A\mathbf{x}$ for a matrix A, that is, if and only if the effect of T on a vector $\mathbf{x}$ can be done by merely multiplying $\mathbf{x}$ by the matrix A.

Theorem 4.10.2 the gives the following (unexpected) **equivalent** condition, that is, this condition is also enough to ensure that T is a linear transformation:

T is a linear transformation from R^n to R^m if and only if, for each $\mathbf{u}, \mathbf{v} \in R^n$ and scalar c :
- $T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$,
- $T(c\mathbf{u}) = cT(\mathbf{u})$.

That this is enough to ensure that T is a linear transformation is not obvious and uses the fact that this condition allows us to find an appropriate matrix A for which $T(\mathbf{x}) = A\mathbf{x}$ (see proof of Theorem 4.10.2).

Now, if every linear transformation $T : R^n \to R^m$ can be "replaced by a matrix", how do we find this matrix? Theorem 4.10.2 shows how:

$$\begin{aligned} \text{Since } T(\mathbf{x}) &= T(x_1\mathbf{e}_1 + x_2\mathbf{e}_2 + ... + x_n\mathbf{e}_n) \\ &= T(x_1\mathbf{e}_1) + T(x_2\mathbf{e}_2) + ... + T(x_n\mathbf{e}_n) \\ &\quad \text{(by the fact that } T(\mathbf{u} + \mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}) \text{ extended to } n \text{ vectors)} \\ &= x_1T(\mathbf{e}_1) + x_2T(\mathbf{e}_2) + ... + x_nT(\mathbf{e}_n) \\ &\quad \text{(by the fact } T(c\mathbf{u}) = cT(\mathbf{u})) \\ &= \begin{bmatrix} & & & & \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & .. & .. & T(\mathbf{e}_n) \\ & & & & \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ . \\ . \\ x_n \end{bmatrix} \end{aligned}$$

since a matrix times a vector is just the linear combination of the columns of the matrix having the components of the vector as coefficients.

This is very important! It says that to find the matrix for T, just "do" T on the basis vectors $\mathbf{e}_1, \mathbf{e}_2, ..., \mathbf{e}_n$ of R^n and make the answers the columns of A, that is,

make

$$\begin{bmatrix} & & & & \\ T(\mathbf{e}_1) & T(\mathbf{e}_2) & .. & .. & T(\mathbf{e}_n) \\ & & & & \end{bmatrix}.$$

Easy! Look at the example following Theorem 4.9.2. This matrix is called the

Standard matrix for T.

5.1 Section 8.1: General linear transformations

Since we deal with general linear transformations in this chapter (not necessarily between R^n and R^m) we use the definition above, but

formulated for general vector spaces V and W. Thus $T : V \to W$ is a linear transformation if, for all vectors $\mathbf{u}, \mathbf{v}$ in V and scalars c :

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u})+T(\mathbf{v}) \\ T(c\mathbf{u}) &= cT(\mathbf{u}). \end{aligned}$$

When $V = W$ we call T a linear *operator*. There is another way to see these conditions. A vector space consists of vectors and implicitly, linear relations between them. So, for example in V we have the vectors $\mathbf{u}$ and $\mathbf{v}$ and the vector $\mathbf{u}+\mathbf{v}$. These vectors have images under T, so $\mathbf{u}$ goes to $T(\mathbf{u})$ and $\mathbf{v}$ goes to $T(\mathbf{v})$ and $\mathbf{u}+\mathbf{v}$ goes to $T(\mathbf{u}+\mathbf{v})$. We want the images of the vectors in V to act the same way as the vectors in V themselves, in some sense we want T to make a "copy" of V in W. For the images to act the same as the vectors in V we would want the vector

$$T(\mathbf{u}+\mathbf{v})$$

to have the same relation to $T(\mathbf{u})$ and $T(\mathbf{v})$ as the vector

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

has to $\mathbf{u}$ and $\mathbf{v}$, namely being

$$\mathbf{u} \text{ plus } \mathbf{v}.$$

So we want

$$T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u})+T(\mathbf{v}).$$

We say that

The image of the sum equals the sum of the images

or that

$$T \text{ "preserves sums".}$$

This is the first part of the definition of linear transformation.

In the same way, we want the relation between $T(\mathbf{u})$ and $T(c\mathbf{u})$ to be the same as that between $\mathbf{u}$ and $c\mathbf{u}$. So we want $T(c\mathbf{u})$ to be c times $T(\mathbf{u})$, hence we want

$$T(c\mathbf{u}) = cT(\mathbf{u}).$$

We say

The image of a scalar multiple equals the scalar multiple of the image

and that

T "preserves scalar multiples".

This idea of "preserving operations" here sum and scalar multiple, is extremely important and will be repeatedly met in your mathematics courses. For example, in abstract algebra we want, for a homomorphism H between groups (the analogue of linear transformation in that context) that

$$H(ab) = H(a)H(b)$$

for exactly analogous reasons. We then say that H *preserves products.*

It can easily be shown that these two conditions are enough to

preserve all linear combinations of V in W

hence the structure of V is reflected in W.

Note that this does not mean that $T(V)$ (the set of all images of elements of V) is literally a copy of V, since T need not be one-to-one. In the extreme case $T(V) = \mathbf{0}$, that is T takes all elements of V to the $\mathbf{0}$ element in W. This is indeed a linear transformation (see Example 2) but clearly T has shrunk V to a single point, losing almost all the structure of V. In general, $T(V)$ will keep some of the structure of V, while identifying different points in V with the same point in W.

To demonstrate or test whether a function $T : V \to W$ is a linear transformation we must test, whether, for all vectors $\mathbf{u}, \mathbf{v}$ in V and scalars c, that

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) \\ T(c\mathbf{u}) &= cT(\mathbf{u}). \end{aligned}$$

To test this you must set $\mathbf{u}$ and $\mathbf{v}$ *general* vectors in V. That is $\mathbf{u}$ and $\mathbf{v}$ will contain variables! So, for example if $V = R^3$ then you could set

$$\begin{aligned} \mathbf{u} &= (u_1, u_2, u_3) \\ \mathbf{v} &= (v_1, v_2, v_3) \end{aligned}$$

which covers all possible $\mathbf{u}, \mathbf{v}$. If $V = P_2$ you could set

$$\begin{aligned} \mathbf{u} &= a_0 + a_1 x + a_2 x^2 \\ \mathbf{v} &= b_0 + b_1 x + b_2 x^2. \end{aligned}$$

Remark: When the book demonstrates that a function T is a linear transformation, it starts with $T(\mathbf{u}+\mathbf{v})$ and then rewrites this to show that it equals $T(\mathbf{u}) + T(\mathbf{v})$. I advise you to calculate $T(\mathbf{u}+\mathbf{v})$ first, then only calculate $T(\mathbf{u}) + T(\mathbf{v})$ and then to check if they are equal. Similarly for showing $T(c\mathbf{u}) = cT(\mathbf{u})$.

So, for example, I recommend doing Example 5 like this:

> Do $T(\mathbf{u}+\mathbf{v})$ first: $T(p_1+p_2) = T(p_1(x)+p_2(x)) = x(p_1(x)+p_2(x))$.
> Now do $T(\mathbf{u})+T(\mathbf{v}) : T(p_1)+T(p_2) = xp_1(x)+xp_2(x)$.
> Now note that $x(p_1(x)+p_2(x)) = xp_1(x)+xp_2(x)$
> (since you can multiply x in).
> Hence T satisfies $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u})+T(\mathbf{v})$.

Do not magically rewrite the one expression into the other, this can be misleading.

Example 1 notes that the previous definition of linear transformation from R^n to R^m is generalized in this definition, and that of course, what qualified as a linear transformation before still does (just take $V = R^n$ and $W = R^m$ in the definition). As was noted in the first paragraph, linear transformations from R^n to R^m can be represented by matrix multiplication so we will sometimes call linear transformations from R^n to R^m *matrix* transformations. Note that of course a matrix transformation $\mathbf{x} \to A\mathbf{x}$ is a linear transformation.

Example 2 shows that the zero transformation is also a linear transformation, even though it "destroys" the entire space V. Since we claimed above that linear transformations preserve the vector space structure of V, what is preserved here? Well $\mathbf{0}$ is at least a vector space, albeit a trivial one, so at least "vector spaceness" is being preserved! In some sense the zero transformation is an extreme case, preserving almost nothing of V.

Example 3 notes that the identity operator is also a linear transformation. This should probably not be a surprise.

Example 4 shows that the dilation (stretch) and contraction (shrink) operators

$$\begin{aligned} T &: V \to V \\ T(\mathbf{v}) &= k\mathbf{v} \end{aligned}$$

are linear.

Example 5 gives a linear transformation T from P_n to P_{n+1}. What T does is take a polynomial and multiply it by $\mathbf{x}$. Hence, e.g.

$$\begin{aligned} T(1+x) &= x(1+2x) \\ &= x+2x^2. \end{aligned}$$

It is interesting that this is indeed a linear transformation.

Example 9 shows that the function which replaces the variable x in a polynomial with $ax+b$ is a linear operator. This is quite unexpected. Do this as an exercise.

Example 6 shows that the function which, given any vector in an inner product space V, takes its inner product with a fixed vector $\mathbf{v}_0$, is a linear transformation.

Example 11 shows that the derivative of continuous functions with continuous first derivatives from $C^1(-\infty,\infty)$ to $F(-\infty,\infty)$ is a linear transformation! Note that the space mapped to is $F(-\infty,\infty)$ (all real-valued functions) and not

$C^1(-\infty, \infty)$ as the derivative of a function with a continuous first derivative need not have a continuous first derivative, but it will still be a continuous function, hence in $F(-\infty, \infty)$. In fact, you could replace F by $C(-\infty, \infty)$, the set of continuous functions on $(-\infty, \infty)$.

The reason that the derivative is a linear transformations is exactly the rules you learnt in your calculus course, namely "the derivative of a sum of functions equals the sum of the separate derivatives" and "the derivative of a scalar multiple of a function equals the scalar multiple times the derivative".

Example 12 shows that integration is also a linear transformation! For similar reasons: "the integral of a sum of functions equals the sum of the separate integrals" and "the integral of a scalar multiple of a function equals the scalar multiple times the integral". You use this in your calculus courses then you "take out" the scalar multiple before you integrate.

Example 7(b) gives a transformation (just another name for a function) which is not linear, namely the determinant mapping. This is because

$$\det(A_1 + A_2) \neq \det A_1 + \det A_2.$$

Think of small A (2×2 say), to illustrate this.

The fact that it took 12 examples to give a function which is not a linear transformation should not mislead you into thinking that "most" functions are. In fact, being a linear transformation is a strong condition which most functions do *not* satisfy.

Properties of linear combinations: As mentioned above, it is enough that

$$\begin{aligned} T(\mathbf{u}+\mathbf{v}) &= T(\mathbf{u}) + T(\mathbf{v}) \\ T(c\mathbf{u}) &= cT(\mathbf{u}) \end{aligned}$$

for T to preserve all linear combinations, that is

$$T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n) = c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + ... + c_nT(\mathbf{v}_n).$$

You can prove this using $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v}), T(c\mathbf{u}) = cT(\mathbf{u})$ and mathematical induction.

Theorem 8.1.1 give three further properties: $T : V \to W$ must map the $\mathbf{0}$ of V to the zero of W, $T(-\mathbf{v}) = -T(\mathbf{v})$ and $T(\mathbf{v} - \mathbf{w}) = T(\mathbf{v}) - T(\mathbf{w})$. The last two just use the fact that

$$-\mathbf{v} = (-1)\mathbf{v}.$$

By the first property, a quick first test to see if a function is a linear transformation is to check whether $T(\mathbf{0}) = \mathbf{0}$. If not, it isn't.

It is enough to know what T does on the *basis* of V to know what T does on *any* vector. Indeed, take T on any vector, say $T(\mathbf{v})$, then, since $\{\mathbf{v}_1, \mathbf{v}_2, ..., \mathbf{v}_n)$ is a basis (say) we have

$$\begin{aligned} & T(\mathbf{v}) \\ = \; & T(c_1\mathbf{v}_1 + c_2\mathbf{v}_2 + ... + c_n\mathbf{v}_n) \\ = \; & c_1T(\mathbf{v}_1) + c_2T(\mathbf{v}_2) + ... + c_nT(\mathbf{v}_n) \\ & \text{by the fact that } T \text{ preserves linear combinations.} \end{aligned}$$

So all we need to know are:

$$T(\mathbf{v}_1), T(\mathbf{v}_2), ..., T(\mathbf{v}_n)$$

to compute the last expression. (all that remains is to multiply by the c's and add!) This is stated by saying that a "linear transformation is completely determined by the images of any set of basis vectors".

Example 10 illustrates this. We are given T on the (non-standard) set of basis vectors for R^3

$$\{(1,1,1),(1,1,0),(1,0,0)\}$$

and asked what T does on a general vector.

To find this, we need to be able to write a general vector in terms of this basis. We first find how to write a general vector (x_1, x_2, x_3) in terms of this basis. (Note that if we knew T on the standard basis it would be easy to find $T(x_1, x_2, x_3)$: Since

$$\begin{aligned} (x_1, x_2, x_3) &= x_1(1,0,0) + x_2(0,1,0) + x_3(0,0,1) \text{ we have} \\ T(x_1, x_2, x_3) &= x_1T(1,0,0) + x_2T(0,1,0) + x_3T(0,0,1).) \end{aligned}$$

but now it is not so easy to write (x_1, x_2, x_3) as a linear combination of

$$\{(1,1,1),(1,1,0),(1,0,0)\}.$$

To write (x_1, x_2, x_3) as a linear combination of $\{(1,1,1),(1,1,0),(1,0,0)\}$ we must find c_1, c_2, c_3 such that

$$(x_1, x_2, x_3) = c_1(1,1,1) + c_2(1,1,0) + c_3(1,0,0).$$

So we must solve this system. (we are implicitly assured that it has exactly one solution, since we are given that it is a basis). It leads to the augmented system

$$\left[\begin{array}{ccc:c} 1 & 1 & 1 & x_1 \\ 1 & 1 & 0 & x_2 \\ 1 & 0 & 0 & x_3 \end{array}\right].$$

Gauss elimination gives

$$\left[\begin{array}{ccc:c} 1 & 0 & 0 & x_3 \\ 0 & 1 & 0 & x_2 - x_3 \\ 0 & 0 & 1 & x_1 - x_2 \end{array}\right].$$

Hence

$$\begin{aligned} c_1 &= x_3 \\ c_2 &= x_2 - x_3 \\ c_3 &= x_1 - x_2. \end{aligned}$$

Hence

$$(x_1, x_2, x_3) = x_3(1,1,1) + (x_2 - x_3)(1,1,0) + (x_1 - x_2)(1,0,0)$$

which you can check by working out the right side.

Now at last

$$\begin{aligned} T(x_1, x_2, x_3) &= T(x_3(1,1,1) + (x_2 - x_3)(1,1,0) + (x_1 - x_2)(1,0,0)) \\ &= x_3 T(1,1,1) + (x_2 - x_3)T(1,1,0) + (x_1 - x_2)T(1,0,0) \\ &= x_3(1,0) + (x_2 - x_3)(2,-1) + (x_1 - x_2)(4,3) \\ &= (4x_1 - 2x_2 - x_3, 3x_1 - 4x_2 + x_3). \end{aligned}$$

Since we are given

$$\begin{aligned} T(1,1,1) &= (1,0) \\ T(1,1,1) &= (2,-1) \\ T(1,0 &= (4,3) \end{aligned}$$

So now we know that T does to any vector. For example

$$T(0,1,0) = (-2,-4).$$

This is a very important example!

Theorem 8.3.1 states that the *composition* $(T_2 \circ T_1)$ of two linear transformations T_1, T_2 is again a linear transformation, where composition is just normal function composition and defined as

$$(T_2 \circ T_1)(\mathbf{u}) = T_2(T_1(\mathbf{u})).$$

That is, do T_1 on $\mathbf{u}$, then do T_2 on the answer.

Make sure you follow the proof. All you are using is that T_1 and T_2 are *themselves* linear transformations.

Example 1 (p.453) gives an example of a composition. Look at it carefully.

Example 2 (p.453) shows that composing with the identity operator leaves T unchanged.

We can of course, compose as many linear transformations as we like.

5.2 Section 8.1: Kernel and Range

The *kernel* of a linear transformation $T : V \to W$ is that set in V that T maps to $\mathbf{0}_W$, the zero in W.

Recall that at least $\mathbf{0}_V$ must be in this set. We denote it by $\ker(T)$.

The set of all vectors in W that are images of at least on vector in V is called the *range* of T, denoted $R(T)$.

Informally, you can think of $R(T)$ as that part of W that T "reaches".

Example 13 points out that the kernel of a matrix multiplication is exactly the null-space of the matrix, indeed

$$\begin{aligned} \mathbf{x} &\in \text{Null-space } A \text{ if and only if} \\ A\mathbf{x} &= \mathbf{0}. \text{ That is } A \text{ maps } \mathbf{x} \text{ to } \mathbf{0}. \end{aligned}$$

Example 14 points out that the kernel and range of the zero transformation are of course V and $\mathbf{0}_W$ respectively. (Since T takes everything to $\mathbf{0}_W$!)

Example 15 points out the kernel and range of the identity operator must be $\mathbf{0}_V$ and V respectively. (T cannot take anything else to $\mathbf{0}$, then it would not be the identity!)

Example 16 has a look at the kernel and range of an Orthogonal projection and Example 17 at a rotation.

Example 18 shows that the kernel of the differentiation transformation is those functions with derivative zero everywhere, that is, the constant functions.

Theorem 8.1.3 proves that the kernel and range of a linear transformation $T : V \to W$ are subspaces! (Of V and W respectively). Recall that to show this we must show them

Non-empty
Closed w.r.t. the operations + and scalar multiple.

This is an important theorem and the proof is not hard. Just an exercise in the definitions. Make sure you understand it.

Recall that the *rank* of a matrix was the dimension of its column(or row) space and the *nullity* the dimension of its null-space. We generalize this to linear transformations T by defining

$$\begin{aligned} &\text{rank}(T) = \text{dimension of } R(T) \\ &\text{nullity}(T) = \text{dimension of } \ker(T). \end{aligned}$$

This subsumes the previous definitions for matrices. (Make sure you know why the dimension of the column space is the same as the dimension of the range, for matrix transformations)

5.3 Section 8.3: Inverse linear transformations

Just as certain functions have inverses, some linear transformations do too. Remember

Linear transformations are just special functions.

The existence of inverses hinges on the fact of T being one-to-one, which means that distinct vectors map to distinct vectors, equivalently, no two different vectors map to the same vector.

In general, we test a transformation for being one-to-one as follows:

Take two elements $\mathbf{u}, \mathbf{v}$ and set $T(\mathbf{u}) = T(\mathbf{v})$. Now see if $T(\mathbf{u}) = T(\mathbf{v})$ forces $\mathbf{u} = \mathbf{v}$. If so, then T is one-to-one. If you can see that there are $\mathbf{u} \neq \mathbf{v}$ with $T(\mathbf{u}) = T(\mathbf{v})$ then T is not one-to-one.

Theorem 8.2.1 Show that the following are equivalent:

$$T \text{ is one-to-one}$$
$$\ker(T) = \{\mathbf{0}\}.$$

It is obvious that it is equivalent to the second condition that $\text{nullity}(T) = 0$.

The theorem is interesting and the proof is easy. Make sure you understand it. This also helps a lot to prove a linear transformation one-to-one as we need only check if $\ker(T) = \{\mathbf{0}\}$.

Example 4 & 5 (p.447) uses the theorem to decide whether linear transformations are one-to-one by looking at the kernel or nullity.

Recall that when we have any function $f : A \to B$ in mathematics for which $R(f) = B$ we say that f is *onto*. So this means that when V is finite dimensional and $T : V \to V$ is one-to-one then we get onto for free.

Example 6 (p.447) gives a linear transformation that is not one-to-one.

Now if T is one-to-one, then every vector in $R(T)$ has exactly one vector mapping to it. This allows us to take the function which starts at the image and maps to the vector which goes to it (called the *pre-image*). Denote this "reverse" function by T^{-1}. Hence

$$T^{-1} : R(T) \to V.$$

Clearly this is a function. But is it a linear transformation? Indeed it is, hence

A one-to-one linear transformation $T : V \to W$ has an inverse $T^{-1} : R(T) \to V$ that is also a linear transformation.

Note that we do not say $T^{-1} : W \to V$ as the notation $f : A \to B$ usually means that f is defined on the entire A, while of course T^{-1} is defined only on $R(T)$ - (you cannot define the pre-image of something that is not an image of anything!)

(You should be mildly surprised that the inverse function must again be a linear transformation.)

Example 1 (p.453) gives the inverse transformation of $T(p) = T(p(x)) = xp(x)$. Not unexpectedly, this is $p(x)/x$.

Example 4 (p.455) checks whether a transformation definable by a matrix, is invertible by checking whether the standard matrix for T is invertible as indeed it is.

Theorem 8.3.2 shows that the composition $T_2 \circ T_1$ of one-to-one linear transformations is again one-to-one and the inverse $(T_2 \circ T_1)^{-1}$ is

$$T_1^{-1} \circ T_2^{-1}$$

Note the implication for matrices.

If $T : V \to W$ is a linear transformation with V and W finite dimensional, and $\dim(W) < \dim(V)$ then T cannot be one-to-one. You can think of V as too "large" to correspond one-to-one with W, analogous to the fact that there is no one-to-one mapping from $\{1, 2\}$ into $\{1\}$. Here though, the "largeness" has to do with the dimension, not the number of elements.

So, for example, there is no one-to-one linear transformation from R^2 to R although there are (surprisingly!) one-to-one functions from R^2 to R(!). (Google it, or see your local library :)

5.4 Section 8.4: Matrices of general linear transformations

Warning: This section is quite demanding, if you understand it on your first reading, you are probably a genius.

As Anton and Rorres state: If $T : V \to W$ is a linear transformation and V and W are finite dimensional, then we can represent T by a matrix! This should be somewhat surprising if you think of some of the linear transformations we have looked at thus far, such as $T : P_2 \to P_3$, defined by

$$T(p(x)) = xp(x).$$

It is not obvious how transformations such as these could be represented as a matrix multiplication

$$A\mathbf{x}.$$

In this very important section we explain how.

Firstly, choose bases for V and W. The key is:

> We will work with the coordinate vectors instead of the vectors themselves.

In this way, we are working in some R^n. For example, instead of working with the polynomials

$$a_0 + a_1x + a_2x^2$$

directly, we choose a basis, say

$$\{1, x, x^2\}$$

and work with the coordinate vector

$$(a_0, a_1, a_2).$$

As example, instead of working with $2 + 3x + 4x^2$ we work with the vector

$$(2, 3, 4)$$

in R^3.

Let B and B' be the chosen bases (of size n and size m respectively) for V and W, then we want a matrix that takes

$$[\mathbf{x}]_B$$

to

$$[T(\mathbf{x})]_{B'}.$$

That is, that takes the coordinate vector of $\mathbf{x}$ to the coordinate vector of $T(\mathbf{x})$.

(Can't we just take one basis? No, since if m and n are different we are already have that basis B cannot be the same as basis B' (the basis have different numbers of vectors!) Also, the bases may be entirely different because what is a convenient basis for V may not be convenient for W.)

This matrix we have in mind must then do the following:

$$A[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}. \qquad (*)$$

So the matrix takes $\mathbf{x}$ in the form of its coordinate vector and outputs $T(\mathbf{x})$ in the form of its coordinate vector. If we like it is of course easy to get back to "the real" $T(\mathbf{x})$ by just taking the linear combination of the basis of W given by $[T(\mathbf{x})]_{B'}$.

That is, if $B' = \{\mathbf{w}_1, \mathbf{w}_2, ..., \mathbf{w}_m\}$, then "the real" $T(\mathbf{x})$ is

$$T(\mathbf{x}) = c_1\mathbf{w}_1 + c_2\mathbf{w}_2 + ... + c_m\mathbf{w}_m$$

where the c_i's are given by $[T(\mathbf{x})]_{B'}$:

$$[T(\mathbf{x})]_{B'} = (c_1, c_2, ..., c_m).$$

Okay, so how do we find this wonderful A? We will use a clever trick to find it. Clearly A must be $m \times n$ and by * we want that

$$A[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}$$

so, in particular this must work for the basis vectors, say $\{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ of B, that is, we want

$$A[\mathbf{u}_1]_B = [T(\mathbf{u}_1)]_{B'}, A[\mathbf{u}_1]_B = [T(\mathbf{u}_1)]_{B'}, ..., A[\mathbf{u}_n]_B = [T(\mathbf{u}_n)]_{B'}$$

But, by definition, the coordinate vector of $\mathbf{u}_1$ with respect to the basis B, namely $[\mathbf{u}_1]_B$ is

$$\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

since

$$\mathbf{u}_1 = 1\mathbf{u}_1 + 0\mathbf{u}_2 ... + 0\mathbf{u}_n$$

similarly for the other $\mathbf{u}_i$.

But then

$$A[\mathbf{u}_1]_B = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix}$$

the first column of A!

Since

$$A[\mathbf{u}_1]_B = [T(\mathbf{u}_1)]_{B'}$$

This means that we have found what the first column of A must be!, namely

$$[T(\mathbf{u}_1)]_{B'}!$$

In other words, the first column of A (remember we are trying to find A) is the image of $\mathbf{u}_1$ under T, written with respect to basis B'. So to find the first column of A we must first find

$$T(\mathbf{u}_1)$$

and then write this vector in terms of the basis for B'.

The same holds for the other columns, hence now we know how to find the matrix for T with respect to the bases B and B'.

Recipe for finding the matrix of a linear operator with respect to bases B and B' :
1) Do T on the basis vectors in B
2) Write each of them in terms of basis B'
3) Make these the columns of the matrix.

This matrix is denoted

$$[T]_{B',B}.$$

Be careful, note the order of the subscripts of T, the order is:

$$\text{"to basis" } (B') \text{ ,"from basis" } (B).$$

In summary then, the matrix we want is

$$[T]_{B',B} = [[T(\mathbf{u}_1)]_{B'} \,|\, [T(\mathbf{u}_2)]_{B'} \,|...|\, [T(\mathbf{u}_n)]_{B'}] \,.$$

And this matrix satisfies:

$$[T]_{B',B}[\mathbf{x}]_B = [T(\mathbf{x})]_{B'}.$$

Don't get scared, this just means that our matrix $[T]_{B',B}$ times $\mathbf{x}$ in its coordinate vector form equals doing T directly on $\mathbf{x}$ itself and then writing the result to base B'.

When the bases are the same, we simply write

$$[T]_B$$

instead of $[T]_{B,B}$.

Example 1 is very important and illustrates this for $T(p(x)) = xp(x)$ the linear operator we were doubtful you could find a matrix for:

Our recipe above says

1) Do T on the basis vectors in B. Okay the basis B is

$$B = \{1, x\}$$

So

$$\begin{aligned} T(1) &= x(1) = x \\ T(2) &= x(x) = x^2. \end{aligned}$$

2) Now write these as linear combinations of the basis vectors in B'. Since $B' = \{1, x, x^2\}$ this is easy, since

$$\begin{aligned} x &= 0(1) + 1(x) + 0(x^2) \\ x^2 &= 0(1) + 0(x) + 1(x^2). \end{aligned}$$

So take the coordinate vectors found here, namely

$$(0, 1, 0)$$
$$(0, 0, 1)$$

and make them the columns of A, so

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

Believe it or not, this matrix will do T's work! For example, take

$$\begin{aligned} \mathbf{u} &= p(x) = 1 + x \text{ so we know} \\ T(\mathbf{u}) &= x + x^2. \end{aligned}$$

> Remember, to find $T(\mathbf{u})$ using A we must first write the vector $\mathbf{u}$ in basis B, and then compute $A[\mathbf{u}]_B$. Depending on the basis B this can be hard work!

To let A do the work of T we must write $\mathbf{u}$ (or $p(x)$) in the B basis so we can compute

$$A[\mathbf{u}]_B$$

Since

$$1 + x = 1(1) + 1(x)$$

we have that

$$[\mathbf{u}]_B = [1 + x]_B = (1, 1)$$

And

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

This last vector is the coordinate vector w.r.t. basis B', hence the vector itself is

$$\begin{aligned} & 0(1) + 1(x) + 1(x^2) \\ = & \ x + x^2 \end{aligned}$$

which is exactly $T(p(x))$!

Example 2 checks that this matrix does the work of T for every vector, basically it does the general case of what we've just done for $1 + x$.

Example 3 illustrates the following point:

If our basis B and/or B' is nasty, we have to work to find $[\mathbf{u}]_B$ and $[T(\mathbf{u}_1)]_{B'}$.

So example 3 gives a linear transformation we must find the matrix for. In this case the bases are not so friendly:

$$\begin{aligned} B &= \left\{\begin{bmatrix} 3 \\ 1 \end{bmatrix}, \begin{bmatrix} 5 \\ 2 \end{bmatrix}\right\} \\ B' &= \left\{\begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}\right\}. \end{aligned}$$

Okay, so we follow our recipe:

1) Compute $T(\mathbf{u}_1), T(\mathbf{u}_2)$ - this is always easy, just apply the definition of T to $\mathbf{u}_1$ and $\mathbf{u}_2$:

$$\begin{aligned} T(\mathbf{u}_1) &= T\begin{bmatrix} 3 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -5(3) + 13(1) \\ -7(3) + 16(1) \end{bmatrix} \\ &= \begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix}. \\ T(\mathbf{u}_1) &= T\begin{bmatrix} 5 \\ 2 \end{bmatrix} = \begin{bmatrix} 2 \\ -5(5) + 13(2) \\ -7(5) + 16(2) \end{bmatrix} \\ &= \begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix}. \end{aligned}$$

2) Write $T(\mathbf{u}_1)$ and $T(\mathbf{u}_2)$ in the basis B'. That means we must find a_1, a_2, a_3 and b_1, b_2, b_3 for which

$$\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + a_3 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \text{ and}$$
$$\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = b_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b_2 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}.$$

So this is work! The first system leads to the augmented matrix:

$$\left[\begin{array}{ccc:c} 1 & -1 & 0 & 1 \\ 0 & 2 & 1 & -2 \\ -1 & 2 & 2 & -5 \end{array}\right]$$

Gauss elimination (you would work here!) gives:

$$\left[\begin{array}{ccc:c} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -2 \end{array}\right].$$

Hence

$$\begin{aligned} a_1 &= 1 \\ a_2 &= 0 \\ a_3 &= -2. \end{aligned}$$

You can check if this satisfies:

$$\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} = a_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + a_2 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + a_3 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$
$$\begin{bmatrix} 1 \\ -2 \\ -5 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 0 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + -2 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

and indeed it does.

Hence $T(\mathbf{u}_1) = 1\mathbf{v}_1 + 0\mathbf{v}_2 - 2\mathbf{v}_3$ and

$$[T(\mathbf{u}_1)]_{B'} = \begin{bmatrix} 1 \\ 0 \\ -2 \end{bmatrix}.$$

The second system leads to the augmented matrix:

$$\left[\begin{array}{ccc:c} 1 & -1 & 0 & 2 \\ 0 & 2 & 1 & 1 \\ -1 & 2 & 2 & -3 \end{array}\right]$$

Gauss elimination (work here!) gives:

$$\left[\begin{array}{ccc:c} 1 & 0 & 0 & 3 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & -1 \end{array}\right].$$

Hence

$$\begin{aligned} b_1 &= 3 \\ b_2 &= 1 \\ b_3 &= -1. \end{aligned}$$

You can check if this satisfies:

$$\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = b_1 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + b_2 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + b_3 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 2 \\ 1 \\ -3 \end{bmatrix} = 3 \begin{bmatrix} 1 \\ 0 \\ -1 \end{bmatrix} + 1 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix} + -1 \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix} \text{ hence}$$

and indeed it does.

Hence $T(\mathbf{u}_2) = 3\mathbf{v}_1 + 1\mathbf{v}_2 - 1\mathbf{v}_3$ and

$$[T(\mathbf{u}_2)]_{B'} = \begin{bmatrix} 3 \\ 1 \\ -1 \end{bmatrix}.$$

3) Make these the columns of A: So with this bit of effort (the effort from finding coordinate vectors for non-nice bases) we get our A

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

The book sweeps all this system solving under the rug with "verify", so be aware that *verify* often means "this is not one step, you need to do work to get this".

Example 4 shows that the matrix for the identity operator is the identity matrix, no matter what the basis is. Is this also the case where we have two different basis for the same space? (Hint: No, why?).

Theorem 8.4.1 states that in the special case where $T : R^n \to R^m$ and B and B' are the standard bases for R^n and R^m respectively, (the (very nice) usual bases $\{(1, 0, .., 0), (0, 1, 0, ..., 0), ..., (0, 0, ..., 1)\}$ of the appropriate size) then $[T]_{B',B}$ is simply $[T]$ which (recall) is the standard matrix for T - the one you find by doing T on the basis vectors and making the results the columns of a matrix. To prove this, imagine going through the standard procedure:

1) Do T on the basis vectors in B
2) Write each of them in terms of basis B'
3) Make these the columns of the matrix.

and see what happens (many steps trivialize).

Example 5 looks at a linear operator on P_2 with luckily again only one basis used. We must

- Find $[T]_B$ with respect to the (nice) basis $B = \{1, x, x^2\}$
- Use $[T]_B$ to find $T(1 + 2x + 3x^2)$
- Find $T(1+2x+3x^2)$ directly and check whether this agrees with our answer above.

Answer

- To find $[T]_B$ we

  1) Do T on the basis vectors in B.
  2) Write each of them in terms of basis B'
  3) Make these the columns of the matrix.

  Luckily $B' = B$ is the nice basis $\{1, x, x^2\}$ so step (2) is immediate.

- To use $[T]_B$ to find $T(1 + 2x + 3x^2)$: we first make a

Remark: To find $T(1 + 2x + 3x^2)$ we must find the coordinate vector of this polynomial w.r.t. the given basis B and then left multiply by $[T]_B$. Only because B is the nice basis $\{1, x, x^2\}$ it this immediate, namely

$$\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

We may as well emphasize this here:

> $[T]_B$ wants to work on vectors to basis B, so if you
> want to use $[T]_B$ to see what T does to some general vector,
> you must first write this vector to basis B!!

It is easy here, because B is such a nice basis, but need not be.
If B was not this nice basis, say $B = \{1, 1 + x, 1 - x^2\}$ and you wanted to find $T(1 + 2x + 3x^2)$ using $[T]_B$ then you would have to find the coordinate vector of $1 + 2x + 3x^2$ with respect to $\{1, 1 + x, 1 - x^2\}$ before multiplying by T_B.

- Okay, back to the example, we need only take

$$[T]_B \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

  Remember, this gives the coordinate vector of $T(1 + 2x + 3x^2)$ w.r.t. basis B. To find $T(1 + 2x + 3x^2)$ "itself" we take the linear combination of vectors in base B prescribed by this coordinate vector.

- To find $T(1 + 2x + 3x^2)$ directly and check whether this agrees with our answer above:

  To find $T(1 + 2x + 3x^2)$ we just apply the definition of T directly to the polynomial $(1 + 2x + 3x^2)$ without going through all the coordinatising. What we get agrees with our result above.

Theorem 8.4.1 gives a formula for finding the matrix of a composition of linear transformations.

Theorem 8.4.2 gives a generalization of the result that a matrix transformation is one-to-one if and only if the matrix itself is invertible: A linear transformation $T : V \to V$ is one-to-one if and only if its matrix $[T]_B$ to basis B is invertible, where B is any basis for V. Further, in this case the matrix for the inverse of T w.r.t. basis B is the inverse of $[T]_B$.

This means we can just take the inverse of $[T]_B$ to find the inverse of T, we need not work with the definition of T directly to find an inverse.

Example 6 uses Theorem 8.4.1 to find the matrix for a composition and checks whether this matrix corresponds to applying the composition directly.

5.5 Section 8.5: Similarity

In the previous section we found matrices for linear transformations $T : V \to W$ when we were given bases B and B' for V and W respectively. Recall that to do this, we

1) Do T on the basis vectors in B
2) Write each of them in terms of basis B'
3) Make these the columns of the matrix.

Clearly, the matrix obtained depends on B and B'. Can we choose B and B' such that the matrix of the linear transformation is really nice, e.g. a diagonal matrix? We can imagine that, ignoring the fact that we may have an awkward basis, this will make our work easier.

This section looks at this.

In the first paragraph the point is made that, for a linear transformation:

Simple bases need not give simple matrices.

W.r.t. the standard basis, the linear operator

$$T\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \begin{bmatrix} x_1 + x_2 \\ -2x_1 + 4x_2 \end{bmatrix}$$

has matrix

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

whereas if we choose the non-standard basis B'

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

the matrix is nicer, namely, the diagonal matrix:

$$T_{B'} = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}.$$

In this section we want to find ways of finding nice matrices for linear transformations, even if we have to take non-nice bases.

If we are going to be changing bases we need to recall how: It follows that if $\mathbf{u}$ is a vector, then the matrix that changes the coordinate vector of $\mathbf{u}$ w.r.t. basis $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_n\}$ to the coordinate vector of $\mathbf{u}$ w.r.t. basis $B = \{\mathbf{u}_1, \mathbf{u}_2, ..., \mathbf{u}_n\}$ is

$$P = \left[\left[\mathbf{u}'_1\right]_B | [\mathbf{u}'_2]_B | ... | \left[\mathbf{u}'_n\right]_B \right].$$

That is, to find P, we write the "from" basis vectors in terms of the "to" basis vectors and make the coordinates the columns. This matrix does what we want, namely

$$P[\mathbf{v}]_{B'} = [\mathbf{v}]_B.$$

Remark: Note that $\mathbf{u}$ is one same vector so, of course, if

$$[\mathbf{u}]_B = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_n \end{bmatrix}$$

and

$$[\mathbf{u}]_{B'} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}$$

then

$$a_1\mathbf{u}_1 + a_2\mathbf{u}_2 + ...a_n\mathbf{u}_n = b_1\mathbf{u}'_1 + b_2\mathbf{u}'_2 + ... + b_n\mathbf{u}'_n = \mathbf{u}.$$

P just changes the b's to the a's.

Theorem 8.5.1 states that the matrix for the identity operator $I : V \to V$ on a vector space with respect to the bases B and B', is just the transition matrix from B to B'. This makes sense: I must leave a vector unchanged, to leave it unchanged it must change its coordinate vector w.r.t. B' to its coordinate vector w.r.t. B but this is exactly what the transition matrix P will do.

We now come to the central problem:

If B and B' are two bases for a finite dimensional vector space V and T is a linear operator, what is the relationship between

$$[T]_B \text{ and } [T]_{B'}?$$

And the answer is (Theorem 8.5.2):

$$[T]_{B'} = P^{-1}[T]_B P$$

where P is the transition matrix from B' to B.

This makes sense: Pretend we want to work to basis B' but we have only $[T]_B$ instead, that is the matrix for when the basis is B. Then we can proceed as follows:

1) rewrite the vector $[\mathbf{u}]_{B'}$ (in our basis B') to the basis $B : P$ does exactly this, so $P([\mathbf{u}]_{B'}) = [\mathbf{u}]_B$

2) Now apply $[T]_B$: Then $[T]_B[\mathbf{u}]_B$ gives the image, but in $B-$coordinates.

3) So transform the resulting vector $[T(\mathbf{u})]_B$ to our basis B' : P^{-1} does exactly this, so $P^{-1}([T]_B[\mathbf{u}]_B)$ gives us what we wanted.

Hence

$$[T]_{B'} = P^{-1}[T]_B P$$

for all $\mathbf{u}$.

Example 1 illustrates this. We go from the matrix

$$[T]_B = \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix}$$

for the standard basis B (check this, we worked this out in a previous example) to the matrix for the basis

$$B' = \left\{ \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 2 \end{bmatrix} \right\}.$$

By the above, to do this, we need the transition matrix P from B' to B so we can get a vector in the form $[T]_B$ wants. Recall that the transition matrix from a basis B' to a basis B is

$$P = \left[\left[\mathbf{u}'_1\right]_B | [\mathbf{u}'_2]_B | ... | \left[\mathbf{u}'_n\right]_B \right].$$

That is, write the basis vectors in B' to basis B:

Now, since

$$\begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 1 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

and

$$\begin{bmatrix} 1 \\ 2 \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 0 \end{bmatrix} + 2 \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

we get

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

We find the inverse of P :

$$P^{-1} = \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix}.$$

Giving

$$\begin{aligned} [T]_{B'} &= P^{-1}[T]_B P \\ &= \begin{bmatrix} 2 & -1 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ -2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}. \end{aligned}$$

So we have found the matrix for the basis B' we want to work with.

In general, when A and B are square matrices and there is an invertible matrix P such that

$$B = P^{-1}AP$$

(such as is the case above) we call A and B *similar*.

Note that this P can be any invertible matrix, it need not be a transition matrix, as above.

A property is called a *similarity invariant* if, when A has the property and B is similar to A, then B must also have the property. Some surprising properties are similarity invariant: Table 1 gives some. (note that the trace - the sum of the diagonal elements is also!)

Since the determinant is a similarity invariant and a change of basis gives a matrix similar to the one for the original basis, it makes sense to define the determinant of a linear operator L as the determinant of the matrix of L with respect to any basis. (Choose any basis B, find $[L]_B$ and then $\det[L]_B$).

Example 2 find the determinant of a linear operator by finding it matrix with respect to the standard basis.

Example 3 is EXTREMELY important. At the beginning of the section we said that, given an operator $T : V \to V$, we want to be able to find a basis for V such that the matrix for T, $[T]_B$ is simple, for example, diagonal. This example illustrates this for a $T : R^3 \to R^3$.

It first finds the standard matrix for T (that is, for $[T]_B$ with B the standard basis for R^3):

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

Now we want to find a basis B' for R^3 for which $[T]$ is diagonal. If by some miracle we find such a diagonal matrix $[T]_{B'}$ then we know that this matrix

$$[T]_{B'} = P^{-1}[T]_B P$$

with $[T]_{B'}$ diagonal.

How on earth does one do this??

Now, out of the blue, the book reminds us that we have previously found a matrix P, namely

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

that diagonalizes our particular $[T]$, in the chapter on eigenvalues when we did diagonalization. So we have a P that will diagonalize $[T]_B$.

But we where asked to find a *basis* with respect to which T has a diagonal matrix. So how does this help?

We now note that we can also see this P as a

transition matrix.

Which does what? Recall that the transition matrix from a basis $B' = \{\mathbf{u}'_1, \mathbf{u}'_2, ..., \mathbf{u}'_n\}$ to a basis B is

$$P = \left[\left[\mathbf{u}'_1\right]_B | [\mathbf{u}'_2]_B | ... | \left[\mathbf{u}'_n\right]_B \right].$$

So what is P the transition matrix for? In particular, we want P to be the transition matrix from base B to base B'. That is, the columns of P are the basis vectors of B' written to basis B. But basis B is just

$$\{(1,0,0),(0,1,0),(0,0,1)\}$$

so the columns of P *are* the basis vectors of B'!

So we have found the basis with respect to which T has a diagonal matrix, namely

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

And where did P come from in the first place? From diagonalizing

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

by building a P with T's eigenvectors as columns!

So the basis with respect to which T has a diagonal matrix, is the basis of eigenvectors of T!

This example contains a lot of information and brings a lot of the work together. Make sure you understand it!

So we can now add to the two equivalent problems from Chapter 7 (section 7.2):

The eigenvector problem:

Given $n \times n$ matrix A, does there exist a basis for R^n consisting of eigenvectors of A?

The matrix diagonalization problem:

Given $n \times n$ matrix A, does there exist an invertible matrix P such that $P^{-1}AP$ is a diagonal matrix?

The diagonal matrix for a linear operator problem:

Given a linear operator T, is there a basis B with respect to which $[T]_B$ is diagonal?