# MAT 2141 - LINEAR ALGEBRA I
# UNIVERSITY OF OTTAWA

ALISTAIR SAVAGE

**Note**: These are notes for the Fall 2010 version of MAT 2141. They will be updated before the start of the Fall 2018 semester.

Contents

**Note to students:** Please let the professor know of any errors you find in these notes (even small typos) so that he can fix them. By doing this, you help him as well as your fellow students.

**Acknowledgement:** Portions of these notes are based on lecture notes by Barry Jessup.

## About the course

*Webpage:* www.mathstat.uottawa.ca/∼asavag2/mat2141

*Professor:* Alistair Savage

*Office Hours:* KED 207G. Mon 10:00–11:00, Wed 11:30–12:30 (subject to change) or by appointment.

*Course text:*

- **Official course text:** Sterling K. Berberian, *Linear Algebra*, Oxford University Press, 1992. This book is out of print. The author and publisher have been kind enough to grant permission for students in this course to purchase photocopies of the text. Students registered in MAT 2141 can purchase a copy of the text for $24.30 (plus tax) at University Centre, Room 0024.
- **Optional text:** Stephen H. Friedberg, Arnold J. Insel, Lawrence E. Spence, *Linear Algebra*, Prentice Hall, 2002. ISBN: 9780130084514. This book is very well suited to this course. It has not been selected as the official course text because of its high price. However, if students wish to purchase it, it is available at the university bookstore (listed as a "recommended text").

*Course Outline:* Will be updated online as the course progresses.

*Assignments:* 6 assignments, due approximately every 2 weeks. Due dates listed on course web page. First one due September 16. Due at the **beginning** of class. **Assignments handed in after class has begun with receive a grade of zero.** Assignments (of more than one page) must be stapled. You should write your name and student number at the top of each page.

*Midterm:* One midterm on Thursday, October 21. Cannot be rescheduled. If you miss it for a legitimate (properly documented) reason, its weight will be added to the weight of the final exam in the computation of your final grade.

*DGDs:* The DGDs will be lead by the professor. We will:

- answer questions from students (concepts students are having difficulty with, etc.),
- supplement material covered in class with extra examples, applications, etc., and
- discuss upcoming assignments, go over difficulties on assignments and midterms, etc.

The DGDs are an important part of the course. Any material covered in the DGDs will be considered "fair game" on exams.

*Computation of grades:* Your grade will be computed in the following way.

- Homework: 20%
- Midterm: 30%
- Final Exam: 50%

In the computation of the homework portion of your grade, the lowest homework mark will be replaced by the grade on the final exam if this is to the student's advantage.

*Virtual Campus:* Grades will be posted here.

*MAT 2141/2541 versus MAT 2342/2742:* MAT 2342 covers similar material to MAT 2141 but is more applied and less abstract and aimed at students studying economics or statistics. Depending on their program, students may have the option of taking either MAT 2141 or MAT 2342. Students in the MAT-ECON program and in statistics are especially encouraged to take MAT 2342. If you're unsure of which course you should be taking, you can discuss your options with the professor.

## Notation

- $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ is the set of nonnegative integers.
- If $Y$ is a subset of a set $X$, then
$$X \setminus Y = \{x \in X \mid x \notin Y\}.$$

**Lecture 1: September 9, 2010**

## 1. Sets with operations (magmas)

**Acknowledgement:** The notes in this section are based on notes by Daniel Daigle.

Up until now, all of the linear algebra you've seen (for instance, in MAT 1341) has involved real or complex numbers. In this course, we will consider a more general (and somewhat more abstract) setting. More precisely, we will allow are scalars to be elements of any *field*. Before giving the definition of a field, we will spend some time discussing sets with operations. We do this in an attempt to eliminate any preconceptions we have about how general operations behave and to make sure that we don't make assumptions which turn out to be false.

### 1.1. Operations on sets.

**Definition 1.1** (Operation on a set)**.** Suppose $E$ is a set. We say that $\star$ is an *operation on* $E$ if for every $x, y \in E$, $x \star y$ is a well-defined element of $E$. A pair $(E, \star)$, where $E$ is a set and $\star$ is an operation on $E$ is called a *set with operation*, or *magma*.

So, loosely speaking, an operation $\star$ on a set $E$ is a "rule" that assigns an element $x \star y$ of $E$ to every pair of elements $(x, y)$ of $E$.

*Remark* 1.2. The term *magma* is not particularly well-known, even among mathematicians. We will use it simply because it is shorter than "set with operation".

**Examples 1.3.**

(a) $(\mathbb{Z}, +)$ is a magma.
(b) $(\mathbb{Z}, -)$ is a magma. However, subtraction is *not* an operation on the set $\mathbb{N} = \{0, 1, 2, 3, 4, \dots\}$ since $x - y$ is not an element of $\mathbb{N}$ for all $x, y \in \mathbb{N}$. Thus, $(\mathbb{N}, -)$ is *not* a magma.
(c) If "$\div$" denotes ordinary division of numbers, then $\div$ is not an operation on $\mathbb{R}$ because $x \div y$ is not defined when $y = 0$. However, if we let $\mathbb{R}^* = \mathbb{R} \setminus \{0\}$, then $(\mathbb{R}^*, \div)$ is a magma.

**Exercise 1.4.** Let $E = \{1, 2, 3\}$ and define $\star$ by the table:

| $\star$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 3 | 4 | 3 |
| 2 | 1 | 1 | 2 |
| 3 | 2 | 2 | 2 |

Is $(E, \star)$ a magma?

**Definition 1.5.** Suppose that $(E, \star)$ is a magma.

(a) We say that $\star$ is *commutative* if $x \star y = y \star x$ for all $x, y \in E$.
(b) We say that $\star$ is *associative* if $(x \star y) \star z = x \star (y \star z)$ for all $x, y, z \in E$.

**Example 1.6.** Consider the set $E = \{1, 2, 3\}$ and the operation $\star$ defined by the table:

| $\star$ | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 3 | 2 | 3 |
| 2 | 1 | 1 | 2 |
| 3 | 2 | 2 | 2 |

The table tells us that $1 \star 2 = 2$ and $2 \star 1 = 1$. Note that $\star$ is indeed an operation on $E$, because it satisfies the requirement of Definition 1.1. So $(E, \star)$ is a magma. Observe that $1 \star 2 \neq 2 \star 1$, so this operation is not commutative.

**Examples 1.7.**

(a) In the magma $(\mathbb{R}, +)$, the operation is commutative and associative.
(b) In the magma $(\mathbb{R}, \cdot)$, the operation is commutative and associative.
(c) In the magma $(\mathbb{R}, -)$, the operation is **not** commutative and **not** associative.
Proof that it is not commutative: $3 - 5 \neq 5 - 3$.
Proof that it is not associative: $(1 - 3) - 5 \neq 1 - (3 - 5)$.

*Remark* 1.8. To show that a magma is **not** associative or commutative, you only need to find **one** counterexample, but to show that a magma **is** associative or commutative, you have to show that the property is satisfied for **all** elements – you **cannot** just give a particular example. For instance, in Example 1.6, we have $2 \star 3 = 2 = 3 \star 2$ but $\star$ is not a commutative operation (as we noted in the example).

1.2. **Use of parentheses.** Suppose $(E, \star)$ is a magma. In expressions like $a \star (b \star (c \star d))$, can we omit the parentheses? The answer depends on whether or not $\star$ is associative.

If $\star$ is an associative operation then we can omit all parentheses. The reason is that it doesn't matter whether by $a \star b \star c$ we mean $(a \star b) \star c$ or $a \star (b \star c)$, since these two quantities are equal. We may also write longer expressions like $a \star b \star c \star d$ without parentheses, because the possible interpretations

$$((a \star b) \star c) \star d, \quad (a \star (b \star c)) \star d, \quad (a \star b) \star (c \star d), \quad a \star ((b \star c) \star d), \quad a \star (b \star (c \star d)),$$

all give the same result, so there is no ambiguity.

The general rule is that parentheses must be used when the operation is not associative. There is one exception to this rule: with the non-associative operation of subtraction, parentheses can be omitted in certain cases. Namely, people have adopted the convention that in expressions such as $a - b - c - d$ (without parentheses), the leftmost operation is evaluated first. In other words, the convention says that $a - b - c - d$ should always be interpreted as meaning $((a - b) - c) - d$. So, if you want to write $((a - b) - c) - d$, then you can omit the parentheses, thanks to this convention; but if you want to write one of

$$(a - (b - c)) - d, \quad (a - b) - (c - d), \quad a - ((b - c) - d), \quad a - (b - (c - d))$$

then the convention does not help you and you must use parentheses.

The convention to evaluate the leftmost operation first (in the absence of parentheses) is universally accepted in the case of subtraction, but not for other operations. If you work with a non-associative operation other than subtraction, you should not assume that you can use that convention.

For instance, let $\wedge$ denote the operation of exponentiation on the set $A = \{1, 2, 3, \dots\}$ of positive integers (i.e., $2 \wedge 3 = 2^3 = 8$ and $3 \wedge 2 = 3^2 = 9$). Then $(A, \wedge)$ is a magma. If you enter the expression $2 \wedge 2 \wedge 3$ on your pocket calculator, it will likely give you the answer 64; so the calculator evaluates the leftmost operation first:

$$2 \wedge 2 \wedge 3 = (2 \wedge 2) \wedge 3 = (2^2)^3 = 4^3 = 64.$$

However, if you ask a mathematician the same question, he will probably do this:

$$2 \wedge 2 \wedge 3 = 2^{2^3} = 2^8 = 256,$$

so the mathematician evaluates the rightmost operation first. For the same reason, a mathematician will always interpret $e^{x^2}$ as meaning $e^{(x^2)}$, never as $(e^x)^2$.

*Conclusion*: use parentheses whenever there is a possibility of confusion.

**Exercise 1.9.** Compute the five possible interpretations of $16 \div 8 \div 4 \div 2$ (two of these interpretations give the same result but the others are all different, so you should get four different numbers).

## 1.3. **Identity elements.**

**Definition 1.10** (Identity element)**.** Suppose that $(E, \star)$ is a magma. If $e$ is an element of $E$ which satisfies

$$e \star x = x = x \star e \quad \text{for all } x \in E,$$

we call $e$ an *identity element* of $(E, \star)$.

**Examples 1.11.**

- 0 is an identity element of $(\mathbb{R}, +)$ because
$$0 + x = x = x + 0 \quad \text{for all } x \in \mathbb{R}.$$
- 1 is an identity element of $(\mathbb{R}, \cdot)$ because
$$1 \cdot x = x = x \cdot 1 \quad \text{for all } x \in \mathbb{R}.$$

*Remark* 1.12. Note that in the above examples, the same set has different identities for different operations. So an identity element depends on the set **and** the operation.

**Example 1.13.** Does $(\mathbb{R}, -)$ have an identity element? Suppose $e$ were an identity element. Then we would have

$$x - e = x \quad \text{for all } x \in \mathbb{R}.$$

This is satisfied for $e = 0$ and so one might be tempted to think that 0 is an identity element. But we must also have

$$e - x = x \quad \text{for all } x \in \mathbb{R}.$$

For each **particular** $x$, the equation $e - x = x$ is only satisfied for $e = 2x$. But $2x$ has a different value for each $x$ (and is only equal to zero when $x = 0$). Therefore, there is no $e \in \mathbb{R}$ that satisfies $e - x = x$ for **all** $x \in \mathbb{R}$. So $(\mathbb{R}, -)$ has no identity element.

**Exercise 1.14.** Verify that the magma $(E, \star)$ in Example 1.6, does not have any identity element.

**Exercise 1.15.** In the magma $(\mathbb{R}^*, \div)$, is the operation commutative? Associative? Does it have an identity element?

**Exercise 1.16.** Consider the set $E = \{1, 2, 3, 4\}$ and the operation $\star$ defined by:

| $\star$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 2 | 1 | 4 |
| 2 | 2 | 3 | 2 | 4 |
| 3 | 1 | 2 | 3 | 4 |
| 4 | 4 | 4 | 4 | 4 |

So $(E, \star)$ is a magma. Does it have an identity element?

We have seen that some magmas do not have an identity element, while others do. Is it possible for a magma to have more than one identity element?

**Theorem 1.17.** *A given magma can have at most one identity element.*

*Proof.* Suppose that $(E, \star)$ is a magma and that $e_1, e_2$ are identity elements of $(E, \star)$. Then

$$\begin{aligned} e_1 &= e_1 \star e_2 \qquad \text{(because } e_2 \text{ is an identity element)} \\ &= e_2 \qquad \text{(because } e_1 \text{ is an identity element).} \end{aligned}$$

□

## 1.4. **Invertible elements.**

**Definition 1.18** (Invertible, inverse)**.** Let $(E, \star)$ be a magma and suppose that $(E, \star)$ has an identity element $e \in E$.

An element $a \in E$ is *invertible* if there exists at least one $x \in E$ satisfying

$$a * x = e = x * a.$$

Any such $x$ is then called *an inverse* of $a$.

*Remark* 1.19. Note that it only makes sense to talk about invertibility and inverses if the magma has an identity element.

**Examples 1.20.**

(a) In $(\mathbb{R}, +)$, is $-2$ invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$(-2) + x = 0 = x + (-2)?$$

Yes, $x = 2$ satisfies this requirement.
So $-2$ is invertible and 2 is an inverse of $-2$.
**Exercise:** Check that, in $(\mathbb{R}, +)$, all elements are invertible.

(b) In $(\mathbb{R}, \cdot)$, is $-2$ invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$(-2) \cdot x = 1 = x \cdot (-2)?$$

Yes, $x = -1/2$ satisfies this requirement.
So $-2$ is invertible and $-1/2$ is an inverse of $-2$.

Again in $(\mathbb{R}, \cdot)$, is 0 invertible? Does there exist $x \in \mathbb{R}$ satisfying

$$0 \cdot x = 1 = x \cdot 0?$$

No, such an $x$ does not exist, so 0 is not invertible. **Exercise:** Check that the set of invertible elements of $(\mathbb{R}, \cdot)$ is equal to $\mathbb{R}^*$.

(c) **Exercise:** Show that 2 is not an invertible element in the magma $(\mathbb{Z}, \cdot)$. Show that the set of invertible elements of $(\mathbb{Z}, \cdot)$ is equal to $\{1, -1\}$.

(d) In $(\mathbb{R}, -)$, is 3 invertible? STOP! This question does not make sense, because $(\mathbb{R}, -)$ does not have an identity element.

When $(E, *)$ does not have an identity element, the concept of invertibility is not defined, so it is absurd to ask whether a given element is invertible. Read Definition 1.18 again and pay attention to the role played by the identity element in that definition. Do you see that, if $e$ does not exist, it makes no sense to speak of invertibility?

(e) Here is a disturbing example. Consider the set $E = \{1, 2, 3, 4\}$ and the operation $*$ defined by the table:

| $*$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 |
| 2 | 2 | 3 | 1 | 1 |
| 3 | 3 | 1 | 3 | 1 |
| 4 | 4 | 1 | 1 | 4 |

So $(E, *)$ is a magma. Observe that $1 * x = x = x * 1$ for all $x \in E$, so $(E, *)$ has an identity element (namely, 1) and consequently the concept of invertibility makes sense.

Is 2 invertible? Does there exist $x \in E$ satisfying

$$2 * x = 1 = x * 2?$$

Yes, there are even two such $x$: $x = 3$ and $x = 4$ satisfy this condition. So 2 is invertible and has two inverses: 3 and 4 are inverses of 2.

**So it is possible for a given element to have more than one inverse!** However, we will see below that this unpleasant phenomenon never happens when the operation is associative.

**Exercise 1.21.** Let $(E, \star)$ be a magma, with identity element $e$. Show that $e$ is invertible, and is its own inverse (more precisely, show that $e$ is the unique inverse of $e$). Let $a, b \in E$; show that if $b$ is an inverse of $a$ then $a$ is an inverse of $b$.

**Exercise 1.22.** Let $\times$ denotes the vector product (or "cross product") on $\mathbb{R}^3$. Then $(\mathbb{R}^3, \times)$ is a magma. Is $(1, 1, 2)$ an invertible element?

## 1.5. **Monoids.**

**Definition 1.23** (Monoid)**.** A *monoid* is a magma $(E, \star)$ such that

- $(E, \star)$ has an identity element, and
- the operation $\star$ is associative.

**Examples 1.24.**

(a) $(\mathbb{R}, +)$, $(\mathbb{R}, \cdot)$, $(\mathbb{Z}, +)$, $(\mathbb{Z}, \cdot)$, $(\mathbb{N}, +)$, $(\mathbb{N}, \cdot)$ are examples of monoids. Also: $(\mathbb{R}^2, +)$, $(\mathbb{R}^3, +)$, $(\mathbb{R}^n, +)$ (for any $n \geq 1$) are monoids. All these examples are commutative monoids.

(b) Let $A = \{x \in \mathbb{Z} \mid x \geq 1\} = \{1, 2, 3, 4, \dots\}$ and let $+$ be ordinary addition. Then $(A, +)$ is a magma but is not a monoid (the operation is associative but there is no identity element).

(c) Consider $(E, *)$ in Example 1.20(e). Then $(E, *)$ is a magma but is not a monoid (there is an identity element but the operation is not associative – **exercise:** prove that $*$ is not associative in that example).

(d) Consider the set $E = \{1, 2, 3, 4\}$ and the operation $*$ defined by the table:

| $*$ | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1 | 2 | 3 | 4 |
| 2 | 2 | 1 | 4 | 3 |
| 3 | 3 | 3 | 3 | 3 |
| 4 | 4 | 4 | 4 | 4 |

So $(E, *)$ is a magma and $(E, *)$ has an identity element (namely, 1). One can also check that $*$ is associative (this takes a bit of work; just take it for granted). So $(E, *)$ is a monoid, and in fact it is a **non-commutative** monoid. **Exercise**: find all invertible elements of this monoid. Also find an inverse of each invertible element.

(e) In MAT 1341, you learned how to multiply matrices. Let $E$ be the set of all $2 \times 2$ matrices with entries in $\mathbb{R}$, and let $\cdot$ denote the multiplication of matrices. Then $\cdot$ is an operation on the set $E$ (because the product of any two $2 \times 2$ matrices with entries in $\mathbb{R}$ is again a $2 \times 2$ matrix with entries in $\mathbb{R}$), and consequently $(E, \cdot)$ is a magma. Moreover,
- the element $\left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$ of the set $E$ is an identity element of this magma;
- in $(E, \cdot)$, the operation is associative (matrix product is associative).

So $(E, \cdot)$ is a monoid; in fact it is a non-commutative monoid. What are the invertible elements in this monoid?

Note that, in any monoid, the concept of invertibility makes sense (because it makes sense in any magma which has an identity element).

**Theorem 1.25.** *In a monoid, a given element can have at most one inverse.*

*Proof.* Let $(E, \star)$ be a monoid and let $e$ be its identity element. Consider $a \in E$ and suppose that $x_1, x_2 \in E$ are inverses of $a$; we have to show that $x_1 = x_2$.

As $x_1, x_2$ are inverses of $a$,

$$a \star x_1 = e = x_1 \star a \quad \text{and} \quad a \star x_2 = e = x_2 \star a.$$

It follows that

$$(*) \qquad x_1 = x_1 \star e = x_1 \star (a \star x_2) = (x_1 \star a) \star x_2 = e \star x_2 = x_2,$$

which proves that $x_1 = x_2$ and hence completes the proof. □

*Remark* 1.26. The five equalities in $(*)$, in the above proof, can be justified as follows:
- the first equality $x_1 = x_1 * e$ is true because $e$ is the identity element;
- the second one is true because $a * x_2 = e$ (because $x_2$ is an inverse of $a$);
- the third by associativity of $*$ (we assumed that $(E, *)$ is a monoid);
- the fourth because $x_1 * a = e$ ($x_1$ is an inverse of $a$);
- the last because $e$ is the identity element.

In mathematics, you are supposed to be able to justify every step of your arguments.

Suppose that $(E, *)$ is a monoid. If $a$ is an invertible element then we know, by Theorem 1.25, that $a$ has exactly one inverse; so it makes sense to speak of **the** inverse of $a$ (as opposed to **an** inverse of $a$).

(a) Suppose that the operation in our monoid is an addition; for instance our monoid could be $(\mathbb{R}, +)$ or $(\mathbb{Z}, +)$ or $(\mathbb{R}^n, +)$, etc. Then we say that our monoid is an *additive monoid* and we often use special notation:
- we write $(E, +)$ instead of $(E, \star)$
- the identity element is *usually* denoted 0
- if $a \in E$ is an invertible element then the inverse of $a$ is *usually* denoted $-a$. If these notations are used then $-a$ is, by definition, the unique element of $E$ satisfying $a + (-a) = 0 = (-a) + a$.

(b) If the operation in our monoid $(E, \star)$ is **not** an addition then the inverse of an element $a$ is often denoted $a^{-1}$. Then, by definition of inverse, $a^{-1}$ is the unique element of $E$ which satisfies

$$a \star a^{-1} = e = a^{-1} \star a,$$

where $e$ denotes the identity element of $(E, \star)$.

**Lecture 2: September 12, 2010**

**Example 1.27** (Functions)**.** If $f$ and $g$ are functions from $\mathbb{R}$ to $\mathbb{R}$, we define a new function "$f + g$" from $\mathbb{R}$ to $\mathbb{R}$ by:

$$(f + g)(x) = f(x) + g(x), \quad \text{for all } x \in \mathbb{R}.$$

For a concrete example of addition of functions, suppose that:

- $f$ is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by $f(x) = \frac{x-1}{x^4+1}$
- $g$ is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by $g(x) = \frac{x+1}{x^4+1}$.

Then the definition of $f + g$ says that $(f + g)(2) = f(2) + g(2) = \frac{1}{17} + \frac{3}{17} = \frac{4}{17}$. More generally, the definition of $f + g$ says that for each $x \in \mathbb{R}$ we have

$$(f + g)(x) = f(x) + g(x) = \frac{x - 1}{x^4 + 1} + \frac{x + 1}{x^4 + 1} = \frac{2x}{x^4 + 1}$$

so $f + g$ is the function from $\mathbb{R}$ to $\mathbb{R}$ defined by

$$(f + g)(x) = \frac{2x}{x^4 + 1}, \quad \text{for all } x \in \mathbb{R}.$$

Let $\mathcal{F}(\mathbb{R})$ denote the set of all functions from $\mathbb{R}$ to $\mathbb{R}$. Then the addition of functions that we just defined is an operation on the set $\mathcal{F}(\mathbb{R})$, i.e., any $f, g \in \mathcal{F}(\mathbb{R})$ determine a well-defined element $f + g \in \mathcal{F}(\mathbb{R})$. So $(\mathcal{F}(\mathbb{R}), +)$ is a magma; let us argue that it is a commutative monoid. Before going into this argument, it is useful to recall what it means for two functions to be equal.

**Definition 1.28** (Equality of functions)**.** Suppose that $f, g$ are functions from $\mathbb{R}$ to $\mathbb{R}$. We say that $f$ and $g$ are equal, and write $f = g$, if for all $x \in \mathbb{R}$, the **real numbers** $f(x)$ and $g(x)$ are equal.

Let us use this definition to show that addition of functions is commutative. Let $f, g$ be functions from $\mathbb{R}$ to $\mathbb{R}$. Then $f + g$ and $g + f$ are functions from $\mathbb{R}$ to $\mathbb{R}$ and, in order to prove that they are equal, we have to show that for each $x \in \mathbb{R}$, the real numbers $(f + g)(x)$ and $(g + f)(x)$ are equal. Now, for any given $x \in \mathbb{R}$, we have

$$(f + g)(x) = f(x) + g(x) = g(x) + f(x) = (g + f)(x)$$

where the equality in the middle is true because addition **of real numbers** is commutative, and the other two equalities are true by definition of addition of functions. This proves that $f + g = g + f$, so addition of functions is a commutative operation on the set $\mathcal{F}(\mathbb{R})$.

**Exercise 1.29.** Imitate the above argument to show that addition of functions is associative.

Let $\mathbf{0}$ denote the function from $\mathbb{R}$ to $\mathbb{R}$ which is identically equal to zero:

$$\mathbf{0}(x) = 0, \quad \text{for all } x \in \mathbb{R}.$$

Then $\mathbf{0} \in \mathcal{F}(\mathbb{R})$.

**Exercise 1.30.** Show that $\mathbf{0} + f = f = f + \mathbf{0}$ for all $f \in \mathcal{F}(\mathbb{R})$. *Remark*: $\mathbf{0} + f$ and $f$ are functions from $\mathbb{R}$ to $\mathbb{R}$; to prove that they are equal, use Definition 1.28.

So $(\mathcal{F}(\mathbb{R}), +)$ has an identity element, namely, the element $\mathbf{0}$ of $\mathcal{F}(\mathbb{R})$. We conclude that $(\mathcal{F}(\mathbb{R}), +)$ is a commutative monoid.

**Exercise 1.31.** Show that, in the monoid $(\mathcal{F}(\mathbb{R}), +)$, each element is invertible. Following the conventions for additive monoids, the inverse of an element $f$ is denoted $-f$. (Here we are talking about the additive inverse of $f$; this has nothing to do with the concept of inverse function.)

**Exercise 1.32.** Define a new operation $\oplus$ on the set $\mathbb{R}$ by $x \oplus y = x + y + 3$, where the "+" in the right hand side is the ordinary addition of numbers. For instance, $7 \oplus 5 = 15$. Then $(\mathbb{R}, \oplus)$ is a magma.

(a) Check that the real number 0 is **not** an identity element of $(\mathbb{R}, \oplus)$.
(b) Find a real number $e$ which is an identity element of $(\mathbb{R}, \oplus)$ (then, by Theorem 1.17, $e$ is the unique identity element of $(\mathbb{R}, \oplus)$). Note that, even though $\oplus$ looks like an addition, it would be a bad idea to denote the identity element of $(\mathbb{R}, \oplus)$ by the symbol 0, because then 0 would denote two different numbers (the real number zero and the identity of the monoid $(\mathbb{R}, \oplus)$). So let $e$ denote the identity element of $(\mathbb{R}, \oplus)$.
(c) Let $x, y, z$ be arbitrary real numbers. Compute the real number $(x \oplus y) \oplus z$; then compute the real number $x \oplus (y \oplus z)$; then check that you obtained the same answer in the two calculations. This argument shows that $(x \oplus y) \oplus z = x \oplus (y \oplus z)$ for all $x, y, z \in \mathbb{R}$, so you have just proved that $\oplus$ is associative. In view of (b), you can conclude that $(\mathbb{R}, \oplus)$ is a monoid. Also prove that $\oplus$ is commutative (compute $x \oplus y$ and $y \oplus x$ separately, and see that you get the same number in the two cases) so $(\mathbb{R}, \oplus)$ is a commutative monoid.
(d) Show that 5 is an invertible element of $(\mathbb{R}, \oplus)$ and find its inverse. (Let us use the notation $\overline{5}$ for the inverse of 5 in $(\mathbb{R}, \oplus)$.)
(e) Show that in the monoid $(\mathbb{R}, \oplus)$, every element is invertible. Find the inverse $\overline{a}$ of each element $a$ of the monoid. What is $\overline{(-3)}$? What is $\overline{0}$?
(f) If we think of $\oplus$ as being an addition, then we would like to define a new operation $\ominus$ on $\mathbb{R}$ which would act like a subtraction. The idea is that subtracting $b$ should be equivalent to adding the inverse of $b$. So we define a new operation $\ominus$ on $\mathbb{R}$ by: given $a, b \in \mathbb{R}$, $a \ominus b = a \oplus \overline{b}$. Compute $0 \ominus 5$ and $5 \ominus 0$ (careful! this is confusing). Note that $(\mathbb{R}, \ominus)$ is a magma, but $\ominus$ is not commutative, not associative, and there is no identity element. Show that the solution $x$ to the equation $x \oplus a = b$ is $x = b \ominus a$ (be careful; show that $x = b \ominus a$ is a solution to the given equation, and show that it is *the only* solution).

**Theorem 1.33.** *If $a, b$ are invertible elements in a monoid $(E, \star)$ then $a \star b$ is invertible and $(a \star b)^{-1} = b^{-1} \star a^{-1}$.*

*Proof.* Let $(E, \star)$ be a monoid, with identity element $e$.

Suppose that $a, b$ are invertible elements of $(E, \star)$. Let $v = b^{-1} \star a^{-1}$ and note that $v$ exists and is an element of $E$. Let us compute $(a \star b) \star v$ and $v \star (a \star b)$; remember that $\star$ is associative, so parentheses can be moved at will, or even omitted:

$$\begin{aligned}(a \star b) \star v &= (a \star b) \star (b^{-1} \star a^{-1}) = a \star (b \star b^{-1}) \star a^{-1} = a \star e \star a^{-1} \\ &= (a \star e) \star a^{-1} = a \star a^{-1} = e, \\ v \star (a \star b) &= (b^{-1} \star a^{-1}) \star (a \star b) = b^{-1} \star (a^{-1} \star a) \star b = b^{-1} \star e \star b \\ &= (b^{-1} \star e) \star b = b^{-1} \star b = e.\end{aligned}$$

So $(a \star b) \star v = e = v \star (a \star b)$; this proves that $a \star b$ is invertible and that its inverse is $v$. $\square$

## 2. Fields

### 2.1. Definition and examples.

**Definition 2.1** (Field)**.** A *field* is a triple $(F, +, \cdot)$ where $F$ is a set, $+$ and $\cdot$ are two operations on $F$, and the following conditions are satisfied:

- Conditions on $(F, +)$:
  - (1) $+$ is commutative and associative,
  - (2) $(F, +)$ has an identity element (which we denote by the symbol 0),
  - (3) every element of $F$ has an additive inverse.
- Conditions on $(F, \cdot)$:
  - (4) $\cdot$ is commutative and associative,
  - (5) $(F, \cdot)$ has an identity element (which we denote by the symbol 1),
  - (6) every element of $F$ other than 0 has a multiplicative inverse.
- Other conditions:
  - (7) *Distributivity*: $a(b + c) = ab + ac$ for all $a, b, c \in F$,
  - (8) $F$ has at least 2 elements.

*Remark* 2.2. We often denote multiplication by juxtaposition. For example, if $a, b \in F$, then $ab$ means $a \cdot b$. We write $F^\times$ for the set of nonzero elements of $F$, that is $F^\times = F \setminus \{0\}$. We also sometimes say "$F$ is a field" (instead of "$(F, +, \cdot)$ is a field") when the two operations (addition and multiplication) are clear from the context.

*Remark* 2.3. Suppose $(F, +, \cdot)$ is a field. The definition implies that $(F, +)$ and $(F, \cdot)$ are both commutative monoids. Recall (Theorem 1.25) that in a monoid, invertible elements have exactly one inverse. Therefore:

- Each $x \in F$ has exactly one additive inverse, which we denote by $-x$. By definition, $-x$ is the unique element of $F$ satisfying $x + (-x) = 0 = (-x) + x$, where 0 is the identity element of $(F, +)$.
- Each $x \in F^\times$ has exactly one multiplicative inverse, which we denote by $x^{-1}$. By definition, $x^{-1}$ is the unique element of $F$ satisfying $xx^{-1} = 1 = x^{-1}x$, where 1 is the identity element of $(F, \cdot)$.

*Remark* 2.4. Note that distributivity and commutativity of multiplication imply

$$(a + b)c = ac + bc \quad \text{for all} \quad a, b, c \in F$$

since

$$(a + b)c = c(a + b) = ca + cb = ac + bc.$$

Thus, we have distributivity in both directions.

**Examples 2.5.**

(a) $\mathbb{C}$, $\mathbb{R}$, and $\mathbb{Q}$ are fields, with the usual addition and multiplication.
(b) $\mathbb{Z}$ (with the usual addition and multiplication) is **not** a field because it contains nonzero elements that do not have multiplicative inverses (for instance, 2).
(c) $\mathbb{R}_{\geq 0} = \{x \in \mathbb{R} \mid x \geq 0\}$ (with the usual addition and multiplication) is **not** a field because it contains elements that do not have additive inverses (for instance, 1).
(d) $M_n(\mathbb{R})$, the set of $n \times n$ matrices (with real number entries) with matrix addition and multiplication, is **not** a field for $n \geq 2$ because there are nonzero matrices (the zero matrix is the additive identity) that do not have multiplicative inverses.

Now let's consider some **new** examples of fields.

**Example 2.6.** (The field $\mathbb{F}_2$) Let $\mathbb{F}_2$ be a set with two elements, which we denote by 0 and 1. Thus, $\mathbb{F}_2 = \{0, 1\}$, where $0 \neq 1$. We define the operations $+$ and $\cdot$ on $F$ by:

| $+$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 1 |
| 1 | 1 | 0 |

and

| $\cdot$ | 0 | 1 |
|---|---|---|
| 0 | 0 | 0 |
| 1 | 0 | 1 |

.

To show that $\mathbb{F}_2$ is a field, we need to show that it satisfies the conditions in Definition 2.1.

- The symmetry of the table defining the operation of addition shows that $+$ is commutative.
- The addition table shows that $0 + 0 = 0$ and $0 + 1 = 1$, hence $0 + x = x$ for all $x \in \mathbb{F}_2$. Since $+$ is commutative, it follows that $x + 0 = x$ for all $x \in \mathbb{F}_2$. Therefore, 0 is an additive identity.
- To prove associativity of $+$, we must show

(*) $$(x + y) + z = x + (y + z) \quad \text{for all} \quad x, y, z \in \mathbb{F}_2.$$

  Since there are two choices for $x$, $y$, and $z$, there are eight equalities to verify. Since 0 is an additive identity, both sides of the equality

  $$(0 + y) + z = 0 + (y + z)$$

  are equal to $y + z$ and thus to each other. Therefore $(*)$ is true when $x = 0$. Similarly, both sides of the equality

  $$(x + y) + 0 = x + (y + 0)$$

  are equal to $x + y$ and thus to each other. Therefore $(*)$ is true when $z = 0$. Also, both sides of the equality

  $$(x + 0) + z = x + (0 + z)$$

  are equal to $x + z$ and thus to each other. Therefore $(*)$ is true when $y = 0$. The only remaining case is when $x = y = z = 1$. Here we check

  $$\left.\begin{array}{c}(1 + 1) + 1 = 0 + 1 = 1 \\ \text{and} \\ 1 + (1 + 1) = 1 + 0 = 1\end{array}\right\} \quad \text{therefore} \quad (1 + 1) + 1 = 1 + (1 + 1).$$

  Thus $(*)$ is true in all cases and hence $+$ is associative.

We have shown that $(\mathbb{F}_2, +)$ satisfies (1), (2) and (3). The proof of (4), (5) and (6) are similar and left as an exercise. It remains to verify (7). First note that the multiplication table shows that

$$0 \cdot x = 0 \quad \text{for all} \quad x \in \mathbb{F}_2.$$

Thus, when $a = 0$, we have

$$a(b + c) = 0(b + c) = 0 = 0 + 0 = 0b + 0c = ab + ac.$$

When $a = 1$, we have

$$a(b + c) = 1(b + c) = b + c = 1b + 1c = ab + ac.$$

Thus (7) is true for all $a, b, c \in \mathbb{F}_2$. Finally, (8) is true since $0 \neq 1$ by definition. Therefore $(\mathbb{F}_2, +, \cdot)$ is a field of two elements.

**Example 2.7** (The field $\mathbb{F}_p$)**.** If $p$ is a prime number, $\mathbb{F}_p = \{\bar{0}, \bar{1}, \bar{2}, \dots, \overline{p-1}\}$ can be made into a field by defining

$$\bar{x} + \bar{y} = \text{remainder after dividing } x + y \text{ by } p,$$
$$\bar{x} \cdot \bar{y} = \text{remainder after dividing } x \cdot y \text{ by } p.$$

For example, in $\mathbb{F}_5$, we have

$$\bar{3} \cdot \bar{4} = \bar{2}, \quad \bar{2} + \bar{4} = \bar{1}, \quad -(\bar{2}) = \bar{3}, \quad \bar{2}^{-1} = \bar{3}.$$

The fields $\mathbb{F}_p$ are called *finite fields*. They are useful in cryptography, coding and other areas of computer science. As long as the context is clear, we sometimes omit the bars and, for instance, write 2 instead of $\bar{2}$ (for example, we did this in Example 2.6).

**Example 2.8** (The field $\mathbb{Q}(\sqrt{2})$)**.** Let

$$\mathbb{Q}(\sqrt{2}) := \{x + y\sqrt{2} \mid x, y \in \mathbb{Q}\}.$$

Then

$$\mathbb{Q} \subsetneq \mathbb{Q}(\sqrt{2}) \subsetneq \mathbb{R}.$$

The fact that $\mathbb{Q} \neq \mathbb{Q}(\sqrt{2})$ follows from the fact that $\sqrt{2}$ is not a rational number (exercise). The fact that $\mathbb{Q}(\sqrt{2}) \neq \mathbb{R}$ follows, for example, from the fact that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$. We claim that $\mathbb{Q}(\sqrt{2})$, with the usual operations of addition and multiplication of real numbers, is a field.

First, we must show that addition and multiplication are actually operations on $\mathbb{Q}(\sqrt{2})$. For $x, y, x', y' \in \mathbb{R}$, we have

$$(x + y\sqrt{2}) + (x' + y'\sqrt{2}) = (x + x') + (y + y')\sqrt{2},$$
$$(x + y\sqrt{2})(x' + y'\sqrt{2}) = (xx' + 2yy') + (xy' + x'y)\sqrt{2}.$$

Thus, addition and multiplication are operations on $\mathbb{Q}(\sqrt{2})$. Conditions (1)–(5), (7), (8) in Definition 2.1 follow easily from the corresponding properties of real numbers (you should go through them and convince yourself that they hold). Condition 6 is not as obvious. We need to show that every nonzero element has a multiplicative inverse. We claim that if $x$ and $y$ are not both zero, then

$$\left(x + y\sqrt{2}\right)^{-1} = \frac{x}{x^2 - 2y^2} - \frac{y}{x^2 - 2y^2}\sqrt{2}.$$

Of course, we must first check that $x^2 - 2y^2 \neq 0$. But if $x^2 - 2y^2 = 0$ then either $x = y = 0$ (which we assumed is not the case), or $\left(\frac{x}{y}\right)^2 = 2$. But since $x, y \in \mathbb{Q}$, $\frac{x}{y}$ is a rational number and since $\sqrt{2}$ is not rational, we cannot have $\left(\frac{x}{y}\right)^2 = 2$. So the expression above is well-defined. Now we check

$$(x + y\sqrt{2})\left(\frac{x}{x^2 - 2y^2} + \frac{-y}{x^2 - 2y^2}\sqrt{2}\right) = \frac{x^2 + 2(-y^2)}{x^2 - 2y^2} + \frac{-xy + yx}{x^2 - 2y^2}\sqrt{2} = \frac{x^2 - 2y^2}{x^2 - 2y^2} = 1,$$

which verifies our claim.

**Examples 2.9.**

(a) $\mathbb{Q}(\sqrt{3})$ is also a field. **Exercise:** Show that $\sqrt{3} \notin \mathbb{Q}(\sqrt{2})$ and so $\mathbb{Q}(\sqrt{2})$ and $\mathbb{Q}(\sqrt{3})$ are different fields.

(b) $\mathbb{Q}(i) = \mathbb{Q}(\sqrt{-1})$ is a field. We have $\mathbb{Q} \subsetneq \mathbb{Q}(i) \subsetneq \mathbb{C}$ and $\mathbb{Q}(i) \not\subseteq \mathbb{R}$.

2.2. **Simplification rules.**

**Proposition 2.10.** *Suppose $a, x, x'$ are elements of a monoid $(M, \star)$ and suppose $a$ is invertible.*

*(a) If $a \star x = a \star x'$, then $x = x'$.*
*(b) If $x \star a = x' \star a$, then $x = x'$.*

*Proof.* (a) Let $e$ be the identity of $(M, \star)$ and let $a^{-1}$ be the inverse of $a$. Suppose that $a \star x = a \star x'$. Then

$$\begin{aligned} & a^{-1} \star (a \star x) = a^{-1} \star (a \star x') \\ \implies & (a^{-1} \star a) \star x = (a^{-1} \star a) \star x' \\ \implies & e \star x = e \star x' \\ \implies & x = x'. \end{aligned}$$

(b) **Exercise.**

□

**Example 2.11.** In the monoid $(\mathbb{R}, \cdot)$, the equality $0 \cdot 2 = 0 \cdot 3$, but $2 \neq 3$. This does not violate Proposition 2.10 because 0 is not an invertible element of $(\mathbb{R}, \cdot)$.

**Example 2.12.** Let $(M_2(\mathbb{R}), \cdot)$ be the monoid of $2 \times 2$ matrices with matrix multiplication. Suppose

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \quad X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \quad X' = \begin{pmatrix} 2 & 1 \\ 0 & 0 \end{pmatrix}.$$

Then

$$AX = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = AX',$$

but $X \neq X'$. Note that $A$ is not an invertible element of $(M_2(\mathbb{R}), \cdot)$.

**Corollary 2.13.** *Let $a, x, x'$ be elements of a field $F$.*

*(a) If $a + x = a + x'$, then $x = x'$.*
*(b) If $ax = ax'$ and $a \neq 0$, then $x = x'$*

*Proof.* (a) Suppose $a + x = a + x'$. Since every element of the monoid $(F, +)$ is invertible (by the definition of a field), $a$ is invertible. Therefore Proposition 2.10 implies $x = x'$.
(b) Suppose $ax = ax'$ and $a \neq 0$. Then $a$ is an invertible element of the monoid $(F, \cdot)$ by the definition of a field, hence Proposition 2.10 implies $x = x'$.

□

**Proposition 2.14.** *Let $F$ be a field.*

*(a) $0x = 0$ for all $x \in F$.*
*(b) The elements 0 and 1 satisfy $1 \neq 0$.*
*(c) $(-1)x = -x$ for all $x \in F$.*
*(d) $(-a)b = -(ab) = a(-b)$ for all $a, b \in F$.*

*Proof.* (a) Let $x \in F$. Then

$$0x + 0x = (0 + 0)x = 0x = 0x + 0.$$

Since $0x$ is invertible in $(F, +)$, we have $0x = 0$ by Proposition 2.10.

(b) We prove this by contradiction. So we suppose $1 = 0$ and show that this implies that one of the axioms of a field is violated. For any element $x \in F$, we have

$$x = 1 \cdot x = 0 \cdot x = 0.$$

But this means that $F$ has only one element, namely 0. Since this contradicts Axiom (8) in Definition 2.1, our assumption that $1 = 0$ must be false. Therefore, $1 \neq 0$.

(c) Let $x \in F$. Then

$$x + (-1)x = 1x + (-1)x = (1 + (-1))x = 0x = 0.$$

This implies that $(-1)x$ is the additive inverse of $x$, hence $(-1)x = -x$.

(d) Let $a, b \in F$. Then

$$(-a)b = ((-1)a)b = (-1)(ab) = -(ab)$$

and

$$a(-b) = a((-1)b) = ((-1)b)a = (-1)(ba) = (-1)(ab) = -(ab).$$

□

**Exercise 2.15.** Show that the element 0 in a field $F$ does not have a multiplicative inverse (*Hint:* Use Proposition 2.14(a) and (b)). Since, by the definition of a field, all nonzero elements have a multiplicative inverse, this shows that the set of elements of a field $F$ with a multiplicative inverse is *exactly* $F^\times$.

**Theorem 2.16.** *Suppose $x, y$ are elements of a field $F$. If $x \neq 0$ and $y \neq 0$, then $xy \neq 0$.*

*Proof.* The proof of this theorem is a problem on the first homework assignment. □

2.3. **Subtraction.**

**Definition 2.17** (Subtraction)**.** Suppose $F$ is a field. We define the operation $-$ of subtraction on $F$ by

$$a - b = a + (-b), \quad \text{for all} \quad a, b \in F.$$

**Exercise 2.18.** Suppose $a, b, c$ are elements of a field $F$. Show that

(a) $a - a = 0$,
(b) $a(b - c) = ab - ac$, and
(c) $(a - b)c = ac - bc$.

2.4. **Preview: linear algebra over fields.** We will see that most of the linear algebra that you saw in MAT 1341 can be done over any field. That is, fields can be the "scalars" in systems of equations and vector spaces. For example, we can solve the system

$$\begin{array}{rcrcl} x_1 & - & 2x_2 & = & 2 \\ x_1 & + & x_2 & = & 0 \end{array}$$

in (a) $\mathbb{R}$, (b) $\mathbb{C}$, (c) $\mathbb{Q}$, (d) $\mathbb{F}_3$.

*Answer:* (a), (b), (c) $x_1 = \frac{2}{3}$, $x_2 = -\frac{2}{3}$. (d) System is inconsistent (that is, it has no solutions).

**Lecture 3: September 16, 2010**

## 3. Vector spaces

Sections 1.1 and 1.2 of the text are review of material covered in MAT 1341. Students should read these sections as a refresher.

### 3.1. Definition and examples.

**Definition 3.1** (Vector space)**.** Let $F$ be a field (whose elements are called *scalars*). A *vector space over* $F$ is

- a set $V$ (whose objects are called *vectors*),
- an operation $+$ on $V$ called *vector addition*, and
- *scalar multiplication:* for each $c \in F$ and $v \in V$, an element $cv \in V$, called the *scalar product* of $c$ and $v$,

such that the following axioms are satisfied:

(a) $(V, +)$ is a commutative monoid (so $+$ is associative, commutative and has an identity element). The identity element is called the *zero vector* and is denoted by $\mathbf{0}$, or $\vec{0}$ (or sometimes just 0, although this can cause confusion with the zero element of the field $F$).
(b) Every vector $v \in V$ has an additive inverse, denoted $-v$.
(c) *Distributivity:* $c(u+v) = cu + cv$ and $(c+d)v = cv + dv$ for all vectors $u, v \in V$ and all scalars $c, d \in F$.
(d) *Associativity:* $(cd)v = c(dv)$ for all vectors $v \in V$ and all scalars $c, d \in F$.
(e) *Unity law:* $1v = v$ for all vectors $v \in V$.

*Remark* 3.2. The text denotes the zero vector by $\theta$.

**Definition 3.3** (Real and complex vector spaces)**.** When $F = \mathbb{R}$ in Definition 3.1, $V$ is called a *real vector space*. When $F = \mathbb{C}$ in Definition 3.1, $V$ is called a *complex vector space*.

**Examples 3.4.**

(a) For each positive integer $n$, $\mathbb{R}^n$ is a real vector space and $\mathbb{C}^n$ is a complex vector space. Here the vector addition and scalar multiplication are the operations you learned in MAT 1341.
(b) Suppose $F$ is a field. For each positive integer $n$,

$$F^n = \{(x_1, \ldots, x_n) \mid x_1, \ldots, x_n \in F\}$$

is a vector space over $F$ with operations defined by

$$(x_1, \ldots, x_n) + (y_1, \ldots, y_n) = (x_1 + y_1, \ldots, x_n + y_n),$$
$$c(x_1, \ldots, x_n) = (cx_1, \ldots, cx_n),$$

for $c, x_1, \ldots, x_n, y_1, \ldots, y_n \in F$. The previous examples of $\mathbb{R}^n$ and $\mathbb{C}^n$ are special cases of this example (where $F = \mathbb{R}$ and $F = \mathbb{C}$).

(c) Suppose $F$ is a field. Then for positive integers $m$ and $n$,

$$M_{m,n}(F) = \{A \mid A \text{ is an } m \times n \text{ matrix with entries in } F\}$$

is a vector space over $F$ with the usual operations of matrix addition and scalar multiplication. Note that matrix multiplication does not play a role when we consider $M_{m,n}(F)$ as a vector space.

(d) In addition to being a complex vector space, $\mathbb{C}$ is *also* a real vector space. **Exercise:** Check that the conditions of Definition 3.1 are satisfied with $V = \mathbb{C}$ and $F = \mathbb{R}$.

**Example 3.5** (Function spaces)**.** Suppose $X$ is a nonempty set and $F$ is a field. Let $\mathcal{F}(X, F)$ or $F^X$ denote the set of functions from $X$ to $F$ (i.e. $F$ valued functions on $X$). When $X = F$, we sometimes write $\mathcal{F}(F)$ in instead of $\mathcal{F}(F, F)$. As in Definition 1.28, for $f, g \in \mathcal{F}(X, F)$, we say $f = g$ if $f(x) = g(x)$ for all $x \in X$. If $f, g \in \mathcal{F}(X, F)$ and $c \in F$, we define functions $f + g, cf \in \mathcal{F}(X, F)$ by

$$(f + g)(x) = f(x) + g(x), \qquad (cf)(x) = cf(x), \qquad \forall\ x \in X.$$

So we define addition and scalar multiplication pointwise. Let $\mathbf{0} \in \mathcal{F}(X, F)$ be the function defined by $\mathbf{0}(x) = 0$ for all $x \in X$ (note the important difference between the zero *function* and the *number* zero). For $f \in \mathcal{F}(X, F)$, define $-f \in \mathcal{F}(X, F)$ by $(-f)(x) = -f(x)$ for all $x \in X$. With these definitions, $\mathcal{F}(X, F)$ is a vector space over $F$.

For example, we can verify the distributivity axiom as follows: For all $f, g \in \mathcal{F}(X, F)$ and $c \in F$, we need to show that $c(f + g) = cf + cg$. These two functions are equal if they are equal at all points of $X$, that is, if

$$(c(f + g))(x) = (cf + cg)(x) \quad \forall\ x \in X.$$

Now, since we know that distributivity holds in the field $F$, for all $x \in X$ we have

$$(c(f+g))(x) = c(f+g)(x) = c(f(x)+g(x)) = cf(x)+cg(x) = (cf)(x)+(cg)(x) = (cf+cg)(x).$$

Thus $c(f + g) = cf + cg$ and so the distributivity axiom in Definition 3.1 holds.
**Exercise:** Verify the remaining axioms of Definition 3.1.

**Example 3.6** ($C^\infty(\mathbb{R})$)**.** Consider the field $F = \mathbb{R}$. Let

$$C^\infty(\mathbb{R}) = \{f \mid f : \mathbb{R} \to \mathbb{R} \text{ has derivatives of all orders}\} \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

As in Example 3.5, we define addition and scalar multiplication pointwise. For example,

$$\sin x,\ \cos x,\ e^x, x^2 + x^3 \in C^\infty(\mathbb{R}),$$

since these functions have derivatives of all orders (they are *infinitely differentiable*). We claim that $C^\infty(\mathbb{R})$ is a vector space over $\mathbb{R}$.

First of all, we must ask ourselves if the operations of addition and scalar multiplication are well-defined on the set $C^\infty(\mathbb{R})$. But we know from calculus that if the $n$th derivatives of $f$ and $g$ exist, then so do the $n$th derivatives of $f+g$ and $cf$, for all $c \in \mathbb{R}$ (we have $(f+g)^{(n)} = f^{(n)}+g^{(n)}$ and $(cf)^{(n)} = cf^{(n)}$). Thus, $C^\infty(\mathbb{R})$ is *closed* under addition and scalar multiplication and so the operations are well-defined.

Next, we must check the axioms of Definition 3.1 are satisfied. Since the zero function $\mathbf{0}$ is infinitely differentiable ($\mathbf{0}' = \mathbf{0}$, $\mathbf{0}'' = \mathbf{0}$, etc.), we have $\mathbf{0} \in C^\infty(\mathbb{R})$. The axioms of Definition 3.1 then hold since they hold in $\mathcal{F}(\mathbb{R}, \mathbb{R})$. Thus $C^\infty(\mathbb{R})$ is a vector space over $\mathbb{R}$.

**Example 3.7.** Let $F = \mathbb{R}$ and

$$V = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = \mathbf{0}\} \subseteq C^\infty(\mathbb{R}) \subseteq \mathcal{F}(\mathbb{R}, \mathbb{R}).$$

For example, if $f(x) = \sin x$, then $f'(x) = \cos x$, $f''(x) = -\sin x$, and so

$$f''(x) + f(x) = -\sin x + \sin x = \mathbf{0},$$

and so $\sin x \in V$. To show that $V$ is a vector space over $\mathbb{R}$, we need to show that it is closed under addition and scalar multiplication (then the rest of the axioms of Definition 3.1 will follow from the fact that $\mathcal{F}(\mathbb{R}, \mathbb{R})$ is a vector space). If $f, g \in V$, then

$$(f+g)'' + (f+g) = f'' + g'' + f + g = (f'' + f) + (g'' + g) = \mathbf{0} + \mathbf{0} = \mathbf{0},$$

and so $f + g \in V$. Thus $V$ is closed under addition. Now, if $f \in V$ and $c \in \mathbb{R}$, we have

$$(cf)'' + cf = cf'' + cf = c(f'' + f) = c\mathbf{0} = \mathbf{0},$$

and so $cf \in V$. Thus $V$ is a vector space over $\mathbb{R}$.

**Example 3.8** (Polynomials)**.** Suppose $F$ is the field $\mathbb{R}$, $\mathbb{C}$ or $\mathbb{Q}$. Let

$$\begin{aligned}\mathcal{P}(F) &= \{p \mid p \text{ is a polynomial with coefficients in } F\} \\ &= \{p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid n \in \mathbb{N},\ a_i \in F\ \forall\ i\}.\end{aligned}$$

Here $t$ is an "indeterminate" (a formal symbol that is manipulated as if it were a number). For $n \in \mathbb{N}$, let

$$\begin{aligned}\mathcal{P}_n(F) &= \{p \in \mathcal{P}(F) \mid p = 0 \text{ or } \deg p \leq n\} \\ &= \{a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid a_i \in F \text{ for all } i\} \subseteq \mathcal{P}(F).\end{aligned}$$

Note the difference between $\mathcal{P}(F)$ and $\mathcal{P}_n(F)$. The elements of $\mathcal{P}(F)$ are polynomials of *any* degree and the elements of $\mathcal{P}_n(F)$ are polynomials of degree at most $n$.

**Exercise 3.9.** Show that $\mathcal{P}(F)$ and $\mathcal{P}_n(F)$ are vector spaces over $F$.

*Remark* 3.10. A polynomial $p(t) = a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0$ defines a function $p : F \to F$. Specifically, $p(t)$ defines the function which maps $c \in F$ to $p(c) = a_n c^n + a_{n-1}c^{n-1} + \cdots + a_0$. There is subtle difference between the *polynomial* $p(t)$ and the *polynomial function* $p$.

**Exercise 3.11.** Fix a positive integer $n$. Why is the set

$$\{a_n t^n + a_{n-1}t^{n-1} + \cdots + a_0 \mid a_i \in F \text{ for all } i,\ a_n \neq 0\}$$

of polynomials of degree exactly $n$ **not** a vector space over $F$?

**Lecture 4: September 20, 2010**

**Example 3.12** (Infinite sequences)**.** Suppose $F$ is a field. Let

$$V = \{(a_1, a_2, \dots) \mid a_i \in F\ \forall\ i\}$$

be the set of all infinite sequence of elements of $F$. We define addition and scalar multiplication componentwise. That is,

$$(a_1, a_2, \dots) + (b_1, b_2, \dots) = (a_1 + b_1, a_2 + b_2, \dots)$$

and

$$c(a_1, a_2, \dots) = (ca_1, ca_2, \dots)$$

for $(a_1, a_2, \dots), (b_1, b_2, \dots) \in V$, $c \in F$. **Exercise:** Show that $V$ is a vector space over $F$.

**Definition 3.13** (Product vector space)**.** Suppose $V_1, V_2, \dots, V_n$ are vector spaces over a field $F$. Define

$$V_1 \times V_2 \times \cdots \times V_n = \{(v_1, v_2, \dots, v_n) \mid v_i \in V_i,\ 1 \leq i \leq n\}.$$

We write $(v_1, v_2, \dots, v_n) = (w_1, w_2, \dots, w_n)$ if $v_i = w_i$ for all $1 \leq i \leq n$. We define addition and scalar multiplication componentwise:

$$(v_1, v_2, \dots, v_n) + (w_1, w_2, \dots, w_n) = (v_1 + w_1, v_2 + w_2, \dots, v_n + w_n),$$
$$c(v_1, v_2, \dots, v_n) = (cv_1, cv_2, \dots, cv_n),$$

for $(v_1, v_2, \dots, v_n), (w_1, w_2, \dots, w_n) \in V_1 \times V_2 \times \cdots \times V_n$ and $c \in F$. We call $V_1 \times V_2 \times \cdots \times V_n$ the *product vector space* of $V_1, V_2, \dots, V_n$.

**Exercise 3.14.** Show that the product vector space is actually a vector space over $F$. That is, show that it satisfies the axioms of Definition 3.1. You will need to use the fact that each $V_i$, $1 \leq i \leq n$, is a vector space over $F$.

## 3.2. Some properties of vector spaces.

**Theorem 3.15.** *Suppose $V$ is a vector space over a field $F$.*

*(a) If $\mathbf{0}' \in V$ is a vector such that $\mathbf{0}' + v = v$ for all $v \in V$, then $\mathbf{0}' = \mathbf{0}$. In other words, the zero vector is unique.*
*(b) If $v + w = \mathbf{0}$ for some $v, w \in V$, then $w = -v$. That is, the additive inverse of a vector is unique (or, negatives of vectors are unique).*
*(c) For all $v \in V$, we have $-(-v) = v$.*
*(d) For all $v \in V$, we have $0v = \mathbf{0}$.*
*(e) For all $c \in F$, $c\mathbf{0} = \mathbf{0}$.*

*Proof.* Here's where our general discussion of magmas and monoids pays off. We know by Definition 3.1 that $(V, +)$ is a commutative monoid. Thus, (a) follows from the uniqueness of identities in magmas and (b) follows from the uniqueness of inverses in monoids. Then (c) is true by applying (b) to the equation $(-v) + v = \mathbf{0}$. To prove (d), note that

$$0v + 0v = (0 + 0)v = 0v = 0v.$$

Adding $-0v$ to both sides then gives

$$0v + 0v + (-0v) = 0v + (-0v) \implies 0v = \mathbf{0},$$

which proves (d). **Exercise:** Prove (e). □

**Corollary 3.16.** *For every vector $v$ and scalar $c$, we have*

$$c(-v) = -(cv) = (-c)v.$$

*Proof.* To prove the first equality, note that

$$c(-v) + cv = c(-v + v) = c\mathbf{0} = \mathbf{0}.$$

Thus $c(-v) = -cv$ by the uniqueness of negatives (Theorem 3.15(b)). Similarly,

$$cv + (-c)v = (c - c)v = 0v = \mathbf{0},$$

and so $(-c)v = -(cv)$ by the uniqueness of negatives. □

**Corollary 3.17.** *For every vector $v$, we have $(-1)v = -v$.*

*Proof.* We have $(-1)v = -(1v) = v$. □

**Theorem 3.18** (The zero vector has no divisors)**.** *Let $V$ be a vector space, $v \in V$, and $c$ a scalar. Then $cv = \mathbf{0}$ if and only if $c = 0$ or $v = \mathbf{0}$.*

*Proof.* We already know from Theorem 3.15 that if $c = 0$ or $v = \mathbf{0}$, then $cv = \mathbf{0}$. So it remains to prove that if $cv = \mathbf{0}$, then either $c = 0$ or $v = \mathbf{0}$. Suppose $cv = \mathbf{0}$. Either $c = 0$ or $c \neq 0$. If $c = 0$, then we're done. So let's consider the remaining case of $c \neq 0$. Then, since $c$ is a nonzero element of a field, it has a multiplicative inverse. Multiplying both sides of the equation $cv = \mathbf{0}$ by $c^{-1}$ gives

$$c^{-1}cv = c^{-1}\mathbf{0} \implies 1v = \mathbf{0} \implies v = \mathbf{0}.$$

□

**Corollary 3.19** (Cancellation laws)**.** *Suppose $u, v$ are vectors and $c, d$ are scalars.*

*(a) If $cu = cv$ and $c \neq 0$, then $u = v$.*
*(b) If $cv = dv$ and $v \neq \mathbf{0}$, then $c = d$.*

*Proof.* (a) Since $c \neq 0$ and all nonzero elements of a field have multiplicative inverses, we can multiply both sides of the equation $cu = cv$ by $c^{-1}$ to get

$$c^{-1}(cu) = c^{-1}cv \implies (c^{-1}c)u = (c^{-1}c)v \implies 1u = 1v \implies u = v.$$

(b) $cv = dv \implies cv + (-dv) = dv + (-dv) \implies (c - d)v = \mathbf{0}$. Then, since $v \neq \mathbf{0}$, we have $c - d = 0$ by Theorem 3.18. Adding $d$ to both sides gives $c - d + d = 0 + d \implies c + 0 = d \implies c = d$.

□

**Definition 3.20** (Subtraction of vectors)**.** For vectors $u, v$, we define

$$u - v = u + (-v).$$

## 3.3. **Linear combinations.**

**Definition 3.21** (Linear combination)**.** Suppose $V$ is a vector space over a field $F$. If $v_1, v_2, \ldots, v_n \in V$ and $c_1, c_2, \ldots, c_n \in F$, then the vector

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n$$

is called a *linear combination* of $v_1, v_2, v_n$. The scalars $c_1, c_2, \ldots, c_n$ are called *coefficients*.

*Remark* 3.22. Note that, in Definition 3.21, we are using the fact that vector addition is commutative when we write sums of multiple (i.e. more than two) vectors without inserting parentheses to tell us in what order we are performing the operations of vector addition.

**Example 3.23.** In the vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$,

$$2\sin x - 3e^x + 5x^3 - 8|x|$$

is a linear combination of the vectors $\sin x$, $e^x$, $x^3$ and $|x|$.

**Example 3.24** (Wave functions in Physics)**.** The theory of vector spaces plays an important role in many areas of physics. For example, in quantum mechanics, the space of wave functions that describe the state of a system of particles is a vector space. The so-called *superposition principle*, that a system can be in a linear combination of states, corresponds to the fact that in vector spaces, one can form linear combinations of vectors.

**Definition 3.25** (Span)**.** Suppose $V$ is a vector space over $F$ and $v_1, v_2, \dots, v_n \in V$. Then

$$\operatorname{Span}\{v_1, v_2, \dots, v_n\} = \{c_1v_1 + c_2v_2 + \cdots + c_nv_n \mid c_1, \dots, c_n \in F\}$$

is the set of all linear combinations of $v_1, v_2, \dots, v_n$ and is call the *span* of this set of vectors. When we wish to emphasize the field we're working with (for instance, when we are working with multiple fields), we write $\operatorname{Span}_F\{v_1, \dots, v_n\}$.

**Example 3.26.** Recall $\mathcal{P}_n(F) = \{a_nt^n + \cdots + a_0 \mid a_i \in F\}$ from Example 3.8. We have

$$\mathcal{P}_n(F) = \operatorname{Span}\{1, t, t^2, \dots, t^n\},$$

where here 1 denotes the constant polynomial function $1(c) = 1$ for all $c \in F$.

**Example 3.27.** Consider the real vector space $\mathcal{F}(\mathbb{R}, \mathbb{R})$. Using the trigonometric sum formulas, we have

$$\cos\left(x + \frac{\pi}{4}\right) = \cos x \cos\frac{\pi}{4} - \sin x \sin\frac{\pi}{4} = \frac{1}{\sqrt{2}}\cos x - \frac{1}{\sqrt{2}}\sin x.$$

Thus $\cos\left(x + \frac{\pi}{4}\right)$ is a linear combination of $\cos x$ and $\sin x$ and so $\cos\left(x + \frac{\pi}{4}\right) \in \operatorname{Span}\{\cos x, \sin x\}$.

**Examples 3.28.**

(a) Consider $\mathbb{C}$ as a vector space over $\mathbb{C}$. We have $\mathbb{C} = \operatorname{Span}_{\mathbb{C}}\{1\}$ since we can write $z = z1$ for any $z \in \mathbb{C}$.
(b) Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$. Then $\mathbb{C} = \operatorname{Span}_{\mathbb{R}}\{1, i\}$. Since we can write *any* complex number as $a + bi$ for some *real* numbers $a, b$.
(c) Consider $\mathbb{R}$ as a vector space over $\mathbb{Q}$. Then $\operatorname{Span}_{\mathbb{Q}}\{1, \sqrt{2}\} = \mathbb{Q}(\sqrt{2})$, the field of Example 2.8.

**Example 3.29.** Consider the vector space $F^n$ over $F$, for some field $F$. Let

$$e_1 = (1, 0, 0, \dots, 0), \quad e_2 = (0, 1, 0, \dots, 0), \quad \dots, \quad e_n = (0, \dots, 0, 0, 1).$$

So for $1 \leq i \leq n$, $e_i$ has a 1 in the $i$th position and a zero everywhere else. Then

$$F^n = \operatorname{Span}_F\{e_1, e_2, \dots, e_n\}$$

since every vector $\mathbf{a} = (a_1, a_2, \dots, a_n) \in F^n$ can be written as

$$\mathbf{a} = a_1e_1 + a_2e_2 + \cdots + a_ne_n = \sum_{k=1}^{n} a_ne_n.$$

Therefore every vector in $F^n$ is a linear combination of the vectors $e_1, e_2, \dots, e_n$.

3.4. **Subspaces.**

**Definition 3.30** (Subspace)**.** Suppose $V$ is a vector space over a field $F$. A subset $U \subseteq V$ is called a *subspace* (or *linear subspace*) of $V$ if

(a) $\mathbf{0} \in U$,
(b) $U$ is closed under vector addition.: $u + v \in U$ for all $u, v \in U$.
(c) $U$ is closed under scalar multiplication: $cu \in U$ for all $u \in U$ and $c \in F$.

**Theorem 3.31** (Subspaces are vector spaces)**.** *If $U$ is a subspace of a vector space $V$, then $U$ is a vector space.*

*Proof.* By Definition 3.30, vector addition and scalar multiplication are well-defined on $U$. Since $\mathbf{0} \in U$ and $-v = (-1)v \in U$ for all $v \in U$, we see that $(U, +)$ is a commutative monoid in which every element has an inverse. The distributivity, associativity and unity law axioms hold since they hold in $V$ (because $V$ is a vector space). □

**Lecture 5: September 23, 2010**

**Examples 3.32.**

(a) If $V$ is a vector space over a field $F$, then $\{0\}$ and $V$ are both subspaces of $V$. These are called the *trivial* subspaces. They are different if $V \neq \{0\}$.
(b) Suppose $A \in M_{m,n}(F)$. Then
$$\ker A = \{v \in F^n \mid Av = 0\}$$
is a subspace of $F^n$. We check the axioms of Definition 3.30. Since $A\mathbf{0} = \mathbf{0}$, we have $\mathbf{0} \in \ker A$. If $u, v \in \ker A$, then $A(u + v) = Au + Av = \mathbf{0} + \mathbf{0} = \mathbf{0}$ and so $u + v \in A$. If $u \in \ker A$ and $c \in F$, then $A(cv) = c(Av) = c\mathbf{0} = \mathbf{0}$ and so $cv \in \ker A$.
(c) Consider $\mathbb{C}$ as a vector space over $\mathbb{R}$. Then $\mathrm{Span}_{\mathbb{R}}\{1\} = \mathbb{R}$ is a subspace of $\mathbb{C}$ over $\mathbb{R}$.
(d) Consider $\mathbb{C}$ as a vector space over $\mathbb{C}$. **Exercise:** Show that the only subspaces are $\{0\}$ and $\mathbb{C}$.
(e) Suppose $F$ is a field and consider $F$ as a vector space over itself. **Exercise:** Show that the only subspaces of $F$ are $\{0\}$ and $F$.
(f) The set
$$V = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f(x) \geq 0 \ \forall \ x \in \mathbb{R}\},$$
is **not** a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$ because it does not contain the additive inverse of every element in $V$. For instance, $f(x) = x^2$ is a function in $V$, but the additive inverse $-f$ is defined by $(-f)(x) = -x^2$ and so $-f$ is not in $V$ (since, for example $(-f)(1) = -1 < 0$). The set $V$ is also not closed under scalar multiplication (**Exercise:** come up with an example to show this).
(g) $C^\infty(\mathbb{R})$ is a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$. The set
$$\{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\}$$
is a subspace of $C^\infty(\mathbb{R})$ *and* a subspace of $\mathcal{F}(\mathbb{R}, \mathbb{R})$.
(h) For any $n \in \mathbb{N}$, $\mathcal{P}_n(F)$ is a subspace of $\mathcal{P}(F)$. When $F = \mathbb{R}$, $\mathcal{P}_n(\mathbb{R})$ is even a subspace of $C^\infty(\mathbb{R})$ since polynomials are infinitely differentiable.

(i) For a field $F$, let
$$W = \{p \in \mathcal{P}(F) \mid p(1) = 0\}.$$
We check that $W$ is a subspace of $\mathcal{P}(F)$. Since $\mathbf{0}(1) = 0$, we have $\mathbf{0} \in W$. Suppose $p, q \in W$, then
$$(p+q)(1) = p(1) + q(1) = 0 + 0 = 0,$$
and so $p + q \in W$. Finally, if $p \in W$ and $c \in F$, then
$$(cp)(1) = cp(1) = c \cdot 0 = 0,$$
and so $cp \in W$. Thus, $W$ is indeed a subspace of $\mathcal{P}$.

(j) For a field $F$, let
$$V = \{p \in \mathcal{P}(F) \mid p(1) = 1\}.$$
Since $\mathbf{0}(1) = 0 \neq 1$, $\mathbf{0} \notin V$ and so $V$ is **not** a subspace of $\mathcal{P}$. The set $V$ also violates the other axioms of a subspace but we can stop here because as soon as a set violates **one** of the axioms, we know it is not a subspace.

**Theorem 3.33.** *If $v_1, v_2, \dots, v_n$ are vectors in a vector space $V$, then $\operatorname{Span}\{v_1, v_2, \dots, v_n\}$ is a subspace of $V$.*

*Proof.* The proof of this theorem is exactly like the proof of the special case where the field is $\mathbb{R}$ (which you saw in MAT 1341). □

**Definition 3.34** (Sum and intersection of subsets)**.** Suppose $M$ and $N$ are subsets of a vector space $V$ (note that they do not need to be subspaces). We define
$$M \cap N = \{v \in V \mid v \in M \text{ and } v \in N\}, \text{ and}$$
$$M + N = \{u + v \mid u \in M,\ v \in N\}.$$
These are called the *intersection* and *sum* (respectively) of $M$ and $N$.

**Theorem 3.35.** *If $U$ and $W$ are **subspaces** of a vector space $V$, then $U \cap W$ and $U + W$ are subspaces as well.*

*Proof.* Let's prove that $U + W$ is a subspace. Since $\mathbf{0} \in U$ and $\mathbf{0} \in W$, we have that
$$\mathbf{0} = \mathbf{0} + \mathbf{0} \in U + W.$$
Next we show that $U + W$ is closed under vector addition. Suppose that $v = u + w$ and $v' = u' + w'$ are two vectors in $U$ (so $u, u' \in U$ and $w, w' \in W$). Then
$$v + v' = (u + w) + (u' + w') = (u + u') + (w + w').$$
Since $U$ is a subspace, it is closed under vector addition and so $u + u' \in U$. Similarly, since $W$ is a subspace, $w + w' \in W$. Thus $v + v' \in U + W$. Finally, we show that $U + W$ is closed under scalar multiplication. Suppose that $v = u + w \in U + W$ (with $u \in U$ and $w \in W$). Then
$$cv = c(u + w) = cu + cw.$$
Since $U$ is a subspace, it is closed under scalar multiplication. Thus $cu \in U$. Similarly, $cw \in W$. Therefore, $cv \in U + W$. So we have shown that $U + W$ is a subspace.
**Exercise:** Show that $U \cap W$ is a subspace of $V$. In fact, it is also a subspace of both $U$ and $W$ (since $U \cap W \subseteq U$ and $U \cap W \subseteq W$). □

*Remark* 3.36. Note that Theorem 3.35 only applies when $U$ and $W$ are **subspaces** of $V$ and not when they are simply subsets.

**Example 3.37.** Suppose $V = F^3$, $U = \{(x, 0, 0) \mid x \in F\}$, $W = \{(0, y, 0) \mid y \in F\}$. **Exercise:** Show that $U$ and $W$ are subspaces of $V$ and $U + W = \{(x, y, 0) \mid x, y \in F\}$ (which is also a subspace). Note that

$$U = \text{Span}\{(1, 0, 0)\}, \quad W = \text{Span}\{(0, 1, 0)\}, \quad U+W = \text{Span}\{(1, 0, 0), (0, 1, 0)\}, \quad U \cap W = \{0\}.$$

**Exercise 3.38.** Suppose $V = F^3$, $U = \{(x, 0, z) \mid x, z \in F\}$, $W = \{(y, y + z, z) \mid y, z \in F\}$.

(a) Show that $U$ and $W$ are subspaces of $V$ and that

$$U = \text{Span}\{(1, 0, 0), (0, 0, 1)\}, \quad W = \text{Span}\{(1, 1, 0), (0, 1, 1)\} = \text{Span}\{(1, 0, -1), (0, 1, 1)\}.$$

(b) Show that $U + W = F^3$.

(c) Show that $U \cap W = \text{Span}\{(1, 0, -1)\}$.

**Corollary 3.39.** *Suppose $U, W$ are subspaces of a vector space $V$ over $F$. Then,*

*(a) $U \cap W$ is the largest subspace of $V$ contained in both $U$ and $W$ (that is, if $X$ is any subspace of $V$ such that $X \subseteq U$ and $X \subseteq W$, then $X \subseteq U \cap W$), and*

*(b) $U + W$ is the smallest subspace of $V$ containing both $U$ and $W$ (that is, if $Y$ is any subspace of $V$ such that $U \subseteq Y$ and $V \subseteq Y$, then $U + W \subseteq Y$).*

*Proof.* (a) Suppose $X$ is a subspace of $V$ such that $X \subseteq U$ and $X \subseteq W$. Then $X \subseteq U \cap W$ by the definition of the intersection $U \cap W$.

(b) Suppose $Y$ is a subspace of $V$ such that $U \subseteq Y$ and $W \subseteq Y$. Let $v \in U + W$. then $v = u + w$ for some $u \in U$ and $w \in W$. Since $u \in U \subseteq Y$, we have $u \in Y$. Similarly, $w \in W \subseteq Y$ implies $w \in Y$. Since $Y$ is a subspace, it is closed under vector addition and so $v = u + w \in Y$. So we have shown that every element of $U + W$ is an element of $Y$. Therefore $U + W \subseteq Y$.

□

**Definition 3.40** (Direct sum)**.** Suppose $V$ is a vector space and $U, W$ are subspaces of $V$ such that

(a) $U \cap W = \{\mathbf{0}\}$, and

(b) $U + W = V$.

We say that $V$ is the *direct sum* of $U$ and $W$ and write $V = U \oplus W$.

**Example 3.41.** Suppose $F$ is a field, $V = F^2$,

$$U = \{(x, 0) \mid x \in F\}, \quad \text{and} \quad W = \{(0, x) \mid x \in F\}.$$

Then $U \cap W = \{\mathbf{0}\}$ and $U + W = V$ since any vector $(x, y) \in V$ can be written as $(x, y) = (x, 0) + (0, y) \in U + W$. Thus $V = U \oplus W$. Sometimes this as written as $F^2 = F \oplus F$, where we identity $U$ and $W$ with $F$ (by considering only the nonzero component).

**Theorem 3.42.** *Suppose $U$ and $W$ are subspace of a vector space $V$. The following statements are equivalent:*

*(a) $V = U \oplus W$,*

*(b) For each $v \in V$, there are **unique** elements $u \in U$ and $v \in V$ such that $v = u + w$.*

*Proof.* We first proof that (a) implies (b). So we assume (a) is true. Suppose $v \in V$. Since $V = U + W$, there exist $u \in U$ and $w \in W$ such that $v = u + w$. Now suppose $v = u' + w'$ for some $u' \in U$ and $w' \in W$. Then

$$0 = v - v = (u + w) - (u' - w') = (u - u') + (w - w') \implies u - u' = w' - w.$$

Now $u - u' \in U$ since $U$ is a subspace (hence closed under vector addition and scalar multiplication). Similarly, $w' - w \in W$. So $u - u' \in U$ and $u - u' = w - w' \in W$. Thus $u - u' \in U \cap W$. But $U \cap W = \{\mathbf{0}\}$. Therefore $u - u' = \mathbf{0}$ and $w - w' = u - u' = \mathbf{0}$. So $u = u'$ and $w = w'$. Hence the representation of $v$ in the form $v = u + w$ is **unique**.
**Exercise:** Prove that (b) implies (a). □

**Exercise 3.43.** Let $U = \text{Span}_{\mathbb{R}}\{(1, 1)\}$ and $V = \text{Span}_{\mathbb{R}}\{1, -1)\}$. Show that $\mathbb{R}^2 = U \oplus V$.

*Remark* 3.44. To show that $S$ is a subspace of a vector space $V$ over a field $F$, it is enough to that $\mathbf{0} \in S$ and

$$u, v \in S,\ c, d \in F \implies cu + dv \in S.$$

Why? Take $c = d = 1$ to see that $S$ is closed under vector addition and then take $d = 0$ to see that $S$ is closed under scalar multiplication.

**Lecture 6: September 27, 2010**

## 4. Linear maps

As a general rule, whenever one introduces a new type of mathematical object (such as vector spaces), it is important to look the natural mappings between them. This helps us understand the relationships between these objects. In the case of vector spaces, the natural mappings between them are linear mappings.

### 4.1. Definition and examples.

**Definition 4.1.** Suppose $V$ and $W$ are vector spaces over the same field $F$. A mapping $T : V \to W$ is said to be *linear* if

(a) $T(v + w) = T(v) + T(w)$ for all $v, w \in V$, and
(b) $T(cv) = cT(v)$ for all $c \in F$ and $v \in V$.

Such a mapping is also called a *linear transformation*.

*Remark* 4.2. If $T : V \to W$ is a linear mapping and $v \in V$, we will sometimes write $Tv$ instead of $T(v)$ (this should remind you of matrix multiplication). We also use the words *map* and *mapping* interchangeably.

*Remark* 4.3. Suppose $V$ and $W$ are vector spaces over a field $F$ and $T : V \to W$ is a linear mapping.

(a) It follows from the definition of a linear mapping that
$$T(cv + dw) = cT(v) + dT(w), \quad \forall\ c, d \in F,\ v, w \in V.$$
(b) In fact, if $c_1, \ldots, c_n \in F$, and $v_1, \ldots, v_n \in V$, then
$$T\left(\sum_{i=1}^{n} c_i v_i\right) = \sum_{i=1}^{n} c_i T(v_i).$$
(c) We have
$$T(\mathbf{0}_V) = T(0 \cdot \mathbf{0}_V) = 0 \cdot T(\mathbf{0}_V) = \mathbf{0}_W.$$
Here we use the notation $\mathbf{0}_V$ and $\mathbf{0}_W$ to distinguish between the zero vectors of $V$ and $W$.
(d) For $v \in V$, we have
$$T(-v) = T((-1)v) = (-1)T(v) = -T(v).$$

Now that we have the definition of a linear map, we turn our attention to constructing some.

**Theorem 4.4.** *Suppose $V$ is a vector space over a field $F$ and $v_1, v_2, \ldots, v_n \in V$ . (Note that we do not require that the $v_i$ be distinct. In other words, some of the $v_i$ may be equal.) Then the mapping $T : F^n \to V$ defined by*
$$T(a_1, a_2, \ldots, a_n) = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n = \sum_{i=1}^{n} a_i v_i.$$
*is linear. Furthermore, it is the unique linear mapping such that $Te_i = v_i$ for all $1 \leq i \leq n$, where the $e_i$ are the vectors described in Example 3.29.*

*Proof.* If $a = (a_1, a_2, \dots, a_n)$ and $b = (b_1, b_2, \dots, b_n)$ are vectors in $F^n$ and $c \in F$, then

$$\begin{aligned} T(a+b) &= T(a_1 + b_1, \dots, a_n + b_n) \\ &= (a_1 + b_1)v_1 + \cdots + (a_n + b_n)v_n \\ &= (a_1 v_1 + \cdots + a_n v_n) + (b_1 v_1 + \cdots + b_n v_n) \\ &= Ta + Tb. \end{aligned}$$

Also,

$$\begin{aligned} T(ca) &= T(ca_1, \dots, ca_n) \\ &= (ca_1)v_1 + \cdots + (ca_n)v_n \\ &= c(a_1 v_1) + \cdots + c(a_n v_n) \\ &= c(a_1 v_1 + \cdots + a_n v_n) \\ &= c(Ta). \end{aligned}$$

Therefore $T$ is linear. Now

$$Te_i = T(0, \dots, 0, 1, 0, \dots, 0) = 0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n = v_i.$$

If $S$ is another linear mapping such that $Se_i = v_i$ for all $i$, then for all $a = (a_1, \dots, a_n) \in F^n$, we have

$$Sa = S(a_1, \dots, a_n) = S(a_1 e_1 + \cdots + a_n e_n) = \sum_{i=1}^{n} a_i S(e_i) = \sum_{i=1}^{n} a_i v_i = Ta.$$

Thus $Sa = Ta$ for all $a \in F^n$ and so $S = T$. □

*Remark* 4.5. Suppose $V$ and $W$ are vector spaces over a field $F$. To show that a map $T : V \to W$ is linear, it is enough to show that

$$T(cu + dv) = cT(u) + dT(v) \ \forall \ c, d \in F, \ u, v \in V.$$

Why? Consider the case $c = d = 1$ and then the case $d = 0$.

**Examples 4.6.**

(a) Let $V = C^\infty(\mathbb{R})$, $F = \mathbb{R}$, and define $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by $D(f) = f'$ (i.e. the map $D$ takes the derivative). Then we know from calculus that

$$D(cf + dg) = (cf + dg)' = cf' + dg' = cD(f) + dD(g),$$

for all $c, d \in \mathbb{R}$ and $f, g \in \mathbb{C}^\infty(\mathbb{R})$.

(b) If

$$C^n(\mathbb{R}) = \{f \in \mathcal{F}(\mathbb{R}, \mathbb{R}) \mid f \text{ is } n \text{ times differentiable}\},$$

then $D : C^n(\mathbb{R}) \to C^{n-1}(\mathbb{R})$ is linear ($n \geq 1$).

(c) Let $V = C^\infty(\mathbb{R})$, $F = \mathbb{R}$, and define $S : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ by

$$(Sf)(x) = \int_0^x f(t)\, dt, \ \forall \ x \in \mathbb{R}.$$

Then we know from calculus that $S(f) \in C^\infty(\mathbb{R})$ and

$$S(cf+dg) = \int_0^x (cf+dg)(t)\,dt = \int_0^x (cf(t)+dg(t))\,dt$$
$$= c\int_0^x f(t)\,dt + d\int_0^x g(t)\,dt = cS(f) + dS(g),$$

for all $c, d \in \mathbb{R}$ and $f, g \in C^\infty(\mathbb{R})$. Thus $S$ is a linear map.

(d) **Exercise:** The map $T : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3$ defined by $T(at^2+bt+c) = (a,b,c)$ is linear.

**Examples 4.7.** The following are linear maps from $\mathbb{R}^3$ to $\mathbb{R}^3$. **Exercise:** Check that the maps are indeed linear.

$$S(x_1,x_2,x_3) = (x_3,x_1,x_2)$$
$$T(x_1,x_2,x_3) = (2x_1-5x_3,0,2x_2)$$

**Example 4.8.** The maps

$$T : \mathbb{R}^3 \to \mathbb{R}^2, \quad T(x_1,x_2,x_3) = (x_2-x_1, 2x_3),$$
$$S : \mathbb{R}^2 \to \mathbb{R}^4, \quad S(x_1,x_2) = (2x_2-x_1, 0, x_1, -4x_2)$$

are linear.

**Example 4.9.** If $V$ is any vector space and $a$ is any scalar, then the mapping $T : V \to V$ defined by $Tv = av$ is linear since for any $u, v \in V$ and scalars $c, d$, we have

$$T(cu+dv) = a(cu+dv) = a(cu)+a(dv) = (ac)u + (ad)v$$
$$= (ca)u + (da)v = c(au)+d(av) = cTu + dTv.$$

Note that we used the commutativity of the field of scalars here.

**Definition 4.10** (Linear form and dual space)**.** Suppose $V$ is a vector space over a field $F$. Then a *linear form* on $V$ is a linear map $V \to F$ and

$$V^* \stackrel{\text{def}}{=} \{f \mid f : V \to F \text{ is linear}\}$$

is called the *dual space* (or simply *dual*) of $V$.

We will soon see that $V^*$ is itself a vector space.

4.2. **Kernel and range.**

**Definition 4.11.** If $f : A \to B$ is any mapping of sets (in particular, $f$ could be a linear mapping between vector spaces) and $A' \subseteq A$, then

$$f(A') = \{f(a) \mid a \in A'\} \subseteq B$$

is called the *image* of $A'$ under $f$. If $B' \subseteq B$, then

$$f^{-1}(B') = \{a \in A \mid f(a) \in B'\} \subseteq A$$

is called the *inverse image* of $B'$ under $f$. Note that we use the notation $f^{-1}(B')$ here even though we do **not** assume that $f$ is invertible. If $B' = \{b\}$ consists of a single element, we will sometimes write $f^{-1}(b)$ for $f^{-1}(\{b\})$ (here one must be extra careful to not confuse this with the inverse function).

**Theorem 4.12.** *Suppose $V$ and $W$ are vector spaces and $T : V \to W$ is a linear mapping.*

(a) *If $M$ is a subspace of $V$, then $T(M)$ is a linear subspace of $W$.*
(b) *If $N$ is a subspace of $W$, then $T^{-1}(N)$ is a subspace of $V$.*

*Proof.* We will use Remark 3.44 in this proof.

(a) Since $M$ is a subspace, we have $\mathbf{0} \in M$ and so $\mathbf{0} = T\mathbf{0} \in T(M)$. Now suppose that $y, y' \in T(M)$ and $c, c'$ are scalars. Then there are $x, x' \in M$ such that $y = Tx$, $y' = Tx'$ and so
$$cy + c'y' = cTx + c'Tx' = T(cx) + T(c'x') = T(cx + c'x') \in T(M)$$
since $cx + c'x' \in M$ (because $M$ is a subspace).
(b) Since $T\mathbf{0} = \mathbf{0} \in N$, we have $\mathbf{0} \in T^{-1}(N)$. Now suppose that $x, x' \in T^{-1}(N)$ and $c, c'$ are scalars. Then
$$T(cx + c'x') = cTx + c'Tx' \in N,$$
since $Tx, Tx' \in N$. Thus $cx + c'x' \in T^{-1}(N)$.

□

**Lecture 7: September 30, 2010**

**Definition 4.13** (Kernel and image)**.** If $T : V \to W$ is a linear map, then
$$\operatorname{Ker} T = T^{-1}(\mathbf{0}) = \{x \mid Tx = \mathbf{0}\}$$
is called the *kernel* (or *null space*) of $T$ and
$$\operatorname{Im} T = T(V) = \{Tx \mid x \in V\}$$
is called the *image* (or *range*) of $T$.

**Corollary 4.14.** *If $T : V \to W$ is a linear mapping, then* $\operatorname{Ker} T$ *is a subspace of $V$ and* $\operatorname{Im} T$ *is a subspace of $W$.*

*Proof.* We take $M = V$ and $N = \{\mathbf{0}\}$ in Theorem 4.12. □

**Example 4.15.** Let $D : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ be the linear mapping given by differentiation. Then
$$\operatorname{Ker} D = \{f \in C^\infty(\mathbb{R}) \mid f \text{ is constant}\},$$
$$\operatorname{Im} D = C^\infty(\mathbb{R}).$$
Why is $\operatorname{Im} T$ all of $C^\infty(\mathbb{R})$? Suppose $f \in C^\infty(\mathbb{R})$. Define $F$ by
$$F(x) = \int_0^x f(t)\,dt.$$
Then we know from calculus that $F$ is differentiable and $DF = F' = f$ (this shows that $F \in C^\infty(\mathbb{R})$). Hence every $f \in C^\infty(\mathbb{R})$ is in the image of $D$.

**Exercise 4.16.** Suppose $S : C^\infty(\mathbb{R}) \to C^\infty(\mathbb{R})$ is the map of Exercise 4.6(c). What is $\operatorname{Im} S$? *Hint:* $S$ is **not** surjective.

**Example 4.17.** Let $f : \mathbb{R}^2 \to \mathbb{R}$ be the linear form defined by $f(x_1, x_2) = x_2 - 3x_1$. Then $\operatorname{Im} f = \mathbb{R}$ and the kernel of $f$ is a line through the origin (the line $x_2 = 3x_1$).

**Example 4.18.** Let $T : \mathbb{R}^2 \to \mathbb{R}^3$ defined by $T(x_1, x_2) = (x_2, 0, x_1)$. Then $\operatorname{Ker} T = \{\mathbf{0}\}$ and the image of $T$ is the '$x, z$-plane' in $\mathbb{R}^3$.

**Theorem 4.19.** *Suppose $T : V \to W$ is a linear mapping. Then $T$ is injective if and only if* $\operatorname{Ker} T = \{\mathbf{0}\}$.

*Proof.* Suppose $\operatorname{Ker} T = \{\mathbf{0}\}$ is injective. Then, for $v, v' \in V$,

$$Tv = Tv' \implies Tv - Tv' = \mathbf{0} \implies T(v - v') = \mathbf{0} \implies v - v' = \mathbf{0} \implies v = v'.$$

Thus $T$ is injective. Now suppose $T$ is injective. Then for $v \in V$,

$$Tv = \mathbf{0} = T\mathbf{0} \implies v = \mathbf{0}.$$

Hence $\operatorname{Ker} T = \{\mathbf{0}\}$. □

*Remark* 4.20. Note that Theorem 4.19 only applies to linear maps. For instance, let $f : \mathbb{R} \to \mathbb{R}$ by the mapping defined by $f(x) = x^2$. Then $f^{-1}(\{0\}) = \{0\}$ but $f$ is **not** injective since, for instance, $f(1) = f(-1)$.

## 4.3. **Vector spaces of linear maps.**

**Definition 4.21.** Suppose $V$ and $W$ are vector spaces over a field $F$. We define

$$\mathcal{L}(V, W) = \{T : V \to W \mid T \text{ is linear}\}.$$

In the case where $V = W$, we write $\mathcal{L}(V)$ for $\mathcal{L}(V, V)$.

**Example 4.22** (Zero linear mapping)**.** For any vector spaces $V$ and $W$ over a field $F$, we have the *zero linear mapping* which maps every element of $V$ to $\mathbf{0} \in W$. We denote the map by 0 (so $0(v) = \mathbf{0}$ for all $v \in V$). Thus $0 \in \mathcal{L}(V, W)$.

**Example 4.23** (Identity mapping)**.** The mapping $I : V \to V$ defined by $Iv = v$ for all $v \in V$ is linear and is called the *identity linear mapping*. We sometimes write $I_V$ when we want to keep track of the vector space on which it acts.

**Example 4.24** (Scalar linear mapping)**.** If $a$ is any scalar, then the mapping $T : V \to V$ defined by $Tv = av$ is linear (see Example 4.9).

**Example 4.25** (Linear forms)**.** The elements of $\mathcal{L}(V, F)$ are precisely the linear forms on $V$.

Recall that in MAT 1341, you learned that every linear map $f$ from $\mathbb{R}^n$ to $\mathbb{R}^m$ (whose elements are written as column vectors) is given by multiplication by the $m \times n$ matrix

$$A = \begin{bmatrix} f(e_1) & f(e_2) & \cdots & f(e_n) \end{bmatrix},$$

where $\{e_1, \ldots, e_n\}$ is the standard basis of $\mathbb{R}^n$ and $f(e_j)$ is the $j$th column of $A$. The matrix $A$ is called the *standard matrix* of the linear map.

**Example 4.26.** The linear map $T : \mathbb{R}^3 \to \mathbb{R}^2$ defined by

$$T(x, y, z) = (y, z)$$

has standard matrix

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

since

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} y \\ z \end{bmatrix} = T \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Later in the course, we'll return to the topic of matrices. We'll see matrices with entries in an arbitrary field $F$ and their relation to linear maps between vectors spaces over $F$.

Our goal now is to turn $\mathcal{L}(V,W)$ into a vector space itself. So we need to define vector addition and scalar multiplication and then check the axioms of a vector space.

**Definition 4.27.** Suppose $V$ and $W$ are vector spaces over a field $F$ and $c \in F$. For $S, T \in \mathcal{L}(V,W)$, define $S+T$ and $cT$ by

$$(S+T)(v) = S(v) + T(v), \quad \forall\ v \in V,$$
$$(cT)(v) = cT(v), \quad \forall\ v \in V.$$

**Lemma 4.28.** *If $S, T \in \mathcal{L}(V,W)$ and $a$ is a scalar, then $S+T,\ aT \in \mathcal{L}(V,W)$.*

*Proof.* We first show that $S+T$ is linear. For $u, v \in V$ and $c, d \in F$, we have

$$\begin{aligned}(S+T)(cu+dv) &= S(cu+dv) + T(cu+dv)\\ &= cS(u) + dS(v) + cT(u) + dT(v)\\ &= cS(u) + cT(u) + dS(v) + dT(v)\\ &= c(S(u)+T(u)) + d(S(v)+T(v))\\ &= c(S+T)(u) + d(S+T)(v).\end{aligned}$$

Therefore, $S+T$ is linear. **Exercise:** Show that $aT$ is linear. □

**Theorem 4.29.** *If $V$ and $W$ are vector spaces over a field $F$, then $\mathcal{L}(V,W)$ is also a vector space over $F$ (with the operations defined in Definition 4.27).*

*Proof.* We must show that the axioms in Definition 3.1 are satisfied.

(a) Vector addition is commutative and associative since for all $S, T, U \in \mathcal{L}(V,W)$ and $v \in V$, we have

$$(S+T)(x) = S(x) + T(x) = T(x) + S(x) = (T+S)(x),$$

and

$$\begin{aligned}((S+T)+U)(x) &= (S+T)(x) + U(x) = S(x) + T(x) + U(x)\\ &= S(x) + (T+U)(x) = (S+(T+U))(x).\end{aligned}$$

The zero linear mapping is an identity for the operation of vector addition since $T + \mathbf{0} = T$ for all $T \in \mathcal{L}(V,W)$.

(b) For all $T \in \mathcal{L}(V,W)$, the mapping $-T = (-1)T$ is linear and is the inverse of $T$ under the operation of vector addition since

$$(T + (-T))(v) = T(v) + (-1)T(v) = \mathbf{0} \quad \forall\ v \in V.$$

**Exercise:** Verify the remaining axioms of Definition 3.1. □

**Example 4.30.** Fix $c \in \mathbb{R}$ and define a map $\varphi_c : \mathcal{F}(\mathbb{R},\mathbb{R}) \to \mathbb{R}$ by

$$\varphi_c(f) = f(c) \quad \forall\ f \in \mathcal{F}(\mathbb{R},\mathbb{R}).$$

We show that $\varphi_c \in \mathcal{F}(\mathbb{R},\mathbb{R})^*$ (see Definition 4.10). To do this, we need to show that $\varphi_c$ is linear. For $k_1, k_2 \in \mathbb{R}$ and $f_1, f_2 \in \mathcal{F}(\mathbb{R},\mathbb{R})$, we have

$$\varphi_c(k_1 f_1 + k_2 f_2) = (k_1 f_1 + k_2 f_2)(c) = k_1 f_1(c) + k_2 f_2(c) = k_1 \varphi_c(f_1) + k_2 \varphi_c(f_2).$$

Thus $\varphi_c$ is linear. This example is important in quantum physics. The so-called 'delta function' which describes the state of a particle located at the point is this type of linear form (i.e. evaluation at the point).

**Definition 4.31** (Composition of mappings)**.** If $A$, $B$ and $C$ are sets (so, for instance, they could be vector spaces), and $T : A \to B$ and $S : B \to C$ are mappings, we write $ST : A \to C$ for the *composite mapping* (or simply the *composition*) defined by

$$(ST)a = S(Ta) \quad \forall a \in A.$$

Composition of mappings is associative, so if $R : C \to D$ is a third mapping, we have $(RS)T = R(ST)$ (**exercise:** check this). We simply write $RST$.

*Remark* 4.32. Sometimes composition mappings are written as $S \circ T$. We use the shorter notation $ST$ to remind us of matrix multiplication. As in MAT 1341, we'll see that composition and matrix multiplication are closely related.

**Lecture 8: October 4, 2010**

**Theorem 4.33.** *The composition of linear maps is a linear map. In other words, if $T \in \mathcal{L}(U, V)$ and $S \in \mathcal{L}(V, W)$, then $ST \in \mathcal{L}(U, W)$.*

*Proof.* For $u, u' \in U$ and $c$ a scalar, we have

$$\begin{aligned}(ST)(u + u') &= S(T(u + u')) \\ &= S(Tu + Tu') \\ &= S(Tu) + S(Tu') \\ &= (ST)(u) + (ST)(u'), \\ (ST)(cu) = S(T(cu)) = S(c(Tu)) &= c(S(Tu)) = c((ST)u).\end{aligned}$$

□

Note that if $U = V = W$, then the above theorem tells us that when $S, T \in \mathcal{L}(V)$, we have $ST \in \mathcal{L}(V)$. Therefore we have **three** operations on $\mathcal{L}(V)$: addition, multiplication by scalars, and composition.

**Definition 4.34** (Powers of a mapping)**.** If $T$ is a map from some set to itself (for instance, if $T \in \mathcal{L}(V)$), then the *powers* of $T$ are defined by

$$T^1 = T, \quad T^2 = TT, \quad T^3 = TT^2, \quad \ldots, \quad T^n = TT^{n-1}, \quad n \geq 2.$$

We also define $T^0$ to be the identity mapping.

4.4. **Isomorphisms.** We will now discuss a precise way in which certain vector spaces are the 'same' (but not necessarily equal).

**Definition 4.35** (Isomorphism)**.** An *isomorphism* is a bijective linear mapping $T : V \to W$, where $V$ and $W$ are vector spaces. We say a vector space $V$ is *isomorphic* to another vector space $W$ (over the same field) if there is an isomorphism $T : V \to W$. We write $V \cong W$ to indicate that $V$ is isomorphic to $W$.

*Remark* 4.36. One should think of isomorphic vector spaces as being 'the same' as far as their vector space properties go. Of course, they make look quite different (and are, in general, not equal). But the isomorphism identifies the two in a way that preserves the operations of a vector space (vector addition and scalar multiplication).

**Example 4.37.** Let

$$V = \mathbb{R}^2,$$
$$W = \{(x, y, 0) \mid x, y \in \mathbb{R}\}.$$

Then $V \cong W$. In order to prove this, we need to find a specific isomorphism. One isomorphism is the map

$$T : V \to W, \quad T(x, y) = (x, y, 0).$$

**Exercise:** Check that the map $T$ is an isomorphism (so you need to show it is linear and bijective).

Note that there is more than one isomorphism from $V$ to $W$. For instance,

$$T_1(x, y) = (y, x, 0) \quad \text{and} \quad T_2(x, y) = (x + y, x - y, 0)$$

are two other isomorphisms from $V$ to $W$.

**Example 4.38.** Consider the map

$$T : \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^3, \quad T(at^2 + bt + c) = (a, b, c).$$

You were asked to show in Exercise 4.6(d) that $T$ is linear. It is easy to see that $T$ is bijective. Thus, $T$ is an isomorphism. Hence $\mathcal{P}_2(\mathbb{R}) \cong \mathbb{R}^3$ (as real vector spaces).

**Exercise 4.39.** Show that $\mathcal{P}_n(\mathbb{R})$ is isomorphic to $\mathbb{R}^{n+1}$.

**Example 4.40.** One can actually show that $\mathcal{P}_n(F) \cong F^{n+1}$ for any **infinite** field $F$. However, things can go wrong if the field is finite. For instance if we work over the field $\mathbb{F}_2$, then the polynomial functions $t$ and $t^2$ are equal! To check this, we just need to check that they take the same value on all of the elements of $\mathbb{F}_2$. Since $0 = 0^2$ and $1 = 1^2$, this verifies that the polynomial functions $t$ and $t^2$ are indeed equal.

**Exercise 4.41.** Let $A = \{1, 2, \dots, n\}$ and $V = \mathcal{F}(A, F)$ for some field $F$. Show $V \cong F^n$ via the bijection $x \mapsto (x(1), x(2), \dots, x(n))$.

*Remark* 4.42. There is a difference between the symbols $\to$ and $\mapsto$. We use $\to$ when we write $T : V \to W$ to indicate that $T$ is a map with domain $V$ and codomain $W$. We use $\mapsto$ to describe a map. So $x \mapsto Tx$ is the map that sends $x$ to $Tx$.

**Example 4.43.** Let $U = \{(x, y, z) \in \mathbb{R}^3 \mid x + y + z = 0\}$. We know from techniques learned in MAT 1341 that $U$ is a subspace of $\mathbb{R}^3$ since it is the solution set to a system of homogeneous equations. In fact, you can use the techniques of MAT 1341 to show that

$$U = \text{Span}\{(-1, 1, 0), (-1, 0, 1)\}$$

Define

$$T : \mathbb{R}^2 \to U, \quad T(x, y) = x(-1, 1, 0) + y(-1, 0, 1) = (-x - y, x, y).$$

Then $T$ is linear since it corresponds to multiplication by the matrix

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

**Exercise:** Show that $T$ is bijective (note that it is **not** bijective as a map $\mathbb{R}^2 \to \mathbb{R}^3$, but it **is** bijective as a map $\mathbb{R}^2 \to U$). So $T$ is an isomorphism and $\mathbb{R}^2 \cong U$.

**Example 4.44.** Remember that $\mathbb{C}$ can be thought of as a real vector space. Then

$$(a, b) \mapsto a + bi, \quad a, b \in \mathbb{R}$$

is an isomorphism from $\mathbb{R}^2$ to $\mathbb{C}$. So $\mathbb{R}^2 \cong \mathbb{C}$ **as real vector spaces**. Not that it is especially important here to state what **type** of isomorphism we have (that is, an isomorphism of real vector spaces). We are not using the complex vector space structure on $\mathbb{C}$.

**Exercise 4.45.** Show that $\mathbb{R}^{2n} \cong \mathbb{C}^n$ as real vector spaces.

Recall that an *inverse* of a map $f : A \to B$ is a map $g : B \to A$ such that $gf$ is the identity map on $A$ and $gf$ is the identity map on $B$. In other words,

$$(gf)(a) = a \ \forall \ a \in A, \quad \text{and} \quad (fg)(b) = b \ \forall \ b \in B.$$

We write $f^{-1}$ for such a map $g$.

Remember that bijective maps have inverses. If $f : A \to B$ is a bijective map from a set $A$ to a set $B$, then we can define the inverse map $f^{-1} : B \to A$, by

$$f^{-1}(b) = a \iff f(a) = b.$$

In other words, for any $b \in B$, $f^{-1}(b)$ is defined to be the unique element $a$ of $A$ such that $f(a) = b$. Such an $a$ exists since $f$ is surjective and it is unique since $f$ is injective.

**Exercise 4.46.**

(a) Show that the inverse of a bijective map is also bijective.
(b) Show that if $f : A \to B$ has an inverse, then $f$ is bijective. Combining this with the above remarks, this means that a map is bijective if and only if it has an inverse.
(c) The composition of two injective maps is itself an injective map.
(d) The composition of two surjective maps is itself a surjective map.
(e) The composition of two bijective maps is itself a bijective map.

**Theorem 4.47.** *Suppose $V$ and $W$ are vector spaces over a field $F$. If $T : V \to W$ is a bijective linear mapping (i.e. an isomorphism), then the inverse mapping $T^{-1} : W \to V$ is also linear (hence is an isomorphism, since it is bijective by Exercise 4.46).*

**Lecture 9: October 7, 2010**

*Proof.* Let $w, w' \in W$ and $c, c' \in F$. We want to show that

$$T^{-1}(cw + c'w') = cT^{-1}w + c'T^{-1}w'. \tag{4.1}$$

Recall that since $T$ is injective, for any $u, v \in V$, we have that $Tu = Tv$ implies $u = v$. So let's apply $T$ to each side of (4.1) and try to show that the results are equal. Applying $T$ to the left-hand side, we get

$$TT^{-1}(cw + c'w') = cw + c'w'.$$

Now, applying $T$ to the right-hand side, we get

$$T(cT^{-1}w + c'T^{-1}w') = cTT^{-1}w + c'TT^{-1}w' = cw + cw',$$

where we have used the fact that $T$ is linear. We thus see that

$$T(T^{-1}(cw + c'w')) = T(cT^{-1}w + c'T^{-1}w')$$

and so $T^{-1}(cw + c'w') = cT^{-1}w + c'T^{-1}w'$ by the injectivity of $T$. □

*Remark* 4.48. By Exercise 4.46, if a linear map $T : V \to W$ has an inverse, then it is an isomorphism. This can be a useful way to show that a give linear map is an isomorphism.

**Example 4.49.** If $A \in M_{n,n}(\mathbb{R})$ is any invertible matrix, then the map $\mathbb{R}^n \to \mathbb{R}^n$ defined by $v \mapsto Av$ for $v \in \mathbb{R}^n$ is an isomorphism. We know this because it is linear (from MAT 1341) and has an inverse, namely the map $\mathbb{R}^n \to \mathbb{R}^n$ given by $v \mapsto A^{-1}v$ for $v \in \mathbb{R}^n$. This is indeed an inverse map since

$$AA^{-1}v = Iv = v \quad \text{and} \quad A^{-1}Av = Iv = v \quad \forall v \in V.$$

We now collect some nice properties of isomorphism.

**Theorem 4.50.** *Suppose $U, V, W$ are vector spaces over the same field. Then*

*(a) $V \cong V$ (isomorphism is* reflexive*),*
*(b) if $V \cong W$, then $W \cong V$ (isomorphism is* symmetric*),*
*(c) if $U \cong V$ and $V \cong W$, then $U \cong W$ (isomorphism is* transitive*).*

*Proof.* (a) The identity map $I : V \to V$ is an isomorphism.

(b) We now know that the inverse of an isomorphism is itself an isomorphism. So suppose $V \cong W$. Then there is an isomorphism $T : V \to W$. Thus, the inverse $T^{-1} : W \to V$ is an isomorphism. Hence $W \cong V$.

(c) If $T : U \to V$ and $S : V \to W$ are isomorphisms, then the composition $ST : U \to W$ is also linear and bijective, hence is an isomorphism.

□

## 4.5. **Equivalence relations and quotient sets.**

**Definition 4.51** (Equivalence relation)**.** Let $X$ be a nonempty set. Suppose that for each ordered pair $(x, y)$ of elements of $X$, we are given a statement $S(x, y)$ about $x$ and $y$. We write $x \sim y$ if the statement $S(x, y)$ is true. We say that $\sim$ is an *equivalence relation* on $X$ if the following three conditions hold:

(a) *reflexivity*: $x \sim x$ for all $x \in X$,
(b) *symmetry*: if $x \sim y$ then $y \sim x$,

(c) *transitivity*: if $x \sim y$ and $y \sim z$, then $x \sim z$.

If $x \sim y$, we say that $x$ is *equivalent* to $y$ (under the relation $\sim$).

**Example 4.52** (Congruence modulo $n$)**.** Fix a positive integer $n$ and consider the set $\mathbb{Z}$ of integers. For $x, y \in \mathbb{Z}$, let $S(x, y)$ be the statement

"$(x - y)$ is an integral multiplie of $n$."

Then $x \sim y$ if and only if $x - y$ is an integrable multiple of $n$. In other words, $x \sim y$ if $(x - y) = kn$ for some $k \in \mathbb{Z}$. If we write $n\mathbb{Z} = \{kn \mid k \in \mathbb{Z}\}$ for the set of integral multiples of $n$, then we have

$$x \sim y \iff (x - y) \in n\mathbb{Z}.$$

Let's check that $\sim$ is an equivalence relation on the set $\mathbb{Z}$.

- *Symmetry*: For all $x \in \mathbb{Z}$, we have $x - x = 0 \in n\mathbb{Z}$, so $x \sim x$.
- *Reflexivity*: Suppose $x \sim y$ for some $x, y \in \mathbb{Z}$. Then $x - y \in n\mathbb{Z}$. But then $y - x = -(x - y) \in n\mathbb{Z}$ (since if $x - y = kz$ for some integer $k$, then $y - x = (-k)z$, which is also an integer multiple of $n$) and so $y \sim x$.
- *Transitivity*: Suppose $x \sim y$ and $y \sim z$ for some $x, y, z \in \mathbb{Z}$. Then $x - y = k_1 n$ for some $k_1 \in \mathbb{Z}$ and $y - z = k_2 n$ for some $k_2 \in \mathbb{Z}$. But then

  $$x - z = (x - y) + (y - z) = k_1 n + k_2 n = (k_1 + k_2)n \in n\mathbb{Z},$$

  since $k_1 + k_2 \in \mathbb{Z}$.

Thus $\sim$ is an equivalence relation on $\mathbb{Z}$. This equivalence relation has its own special notation. If $x \sim y$, we write

$$x \equiv y \mod n,$$

and say $x$ is *congruent to $y$ modulo $n$*.

**Example 4.53.** Suppose $A$ and $B$ are sets and $f : A \to B$ is any function (map of sets). For $x, y \in A$, write $x \sim y$ if $f(x) = f(y)$. Let's check that $\sim$ is an equivalence relation.

(a) *Reflexive:* For all $x \in A$, we have $x \sim x$ since $f(x) = f(x)$.
(b) *Symmetric:* If $x, y \in A$ such that $x \sim y$, then $f(x) = f(y)$. Hence $f(y) = f(x)$ and so $y \sim x$.
(c) *Transitive:* Suppose $x, y, z \in A$ such that $x \sim y$ and $y \sim z$. Then $f(x) = f(y) = f(z)$ and so $x \sim z$.

**Example 4.54.** Suppose $M$ is a subspace of a vector space $V$. For $u, v \in V$, write $u \sim v$ if $u - v \in M$. Then $\sim$ is an equivalence relations on $V$. We check the three properties of an equivalence relation.

(a) *Reflexive:* For all $v \in V$, we have $v - v = \mathbf{0} \in M$, since $M$ is a subspace.
(b) *Symmetric:* For all $u, v \in V$ such that $u \sim v$, we have $u - v \in M$. Then

  $$v - u = -(u - v) = (-1)(u - v) \in M$$

  since $M$ is a subspace and hence closed under scalar multiplication.
(c) *Transitive:* Suppose $u, v, w \in V$ such that $u \sim v$ and $v \sim w$. Then $u - v \in M$ and $v - w \in M$. Then

  $$u - w = (u - v) + (v - w) \in M$$

  since $M$ is a subspace and hence closed under vector addition.

**Lecture 10: October 14, 2010**

**Definition 4.55** (Equivalence class)**.** Suppose $\sim$ is an equivalence relation on a set $X$. Then, for $x \in X$, we define

$$[x] = \{y \in X \mid x \sim y\}$$

to be the set of all elements of $X$ that are equivalent to $x$. We call $[x]$ the *equivalence class* of $x$.

**Example 4.56.** Consider the equivalence relation on $\mathbb{Z}$ given by congruence modulo $n$ (Example 4.52). Then, for $a \in \mathbb{Z}$, we have

$$[a] = \{b \in \mathbb{Z} \mid a - b \in n\mathbb{Z}\} = \{a + kn \mid k \in \mathbb{Z}\} \stackrel{\text{def}}{=} a + n\mathbb{Z}.$$

**Example 4.57.** Suppose $A, B$ are sets, $f : A \to B$ is any function, and $\sim$ is the equivalence relation of Example 4.53. Then, for $x \in A$,

$$[x] = \{y \in A \mid f(y) = f(x)\} = f^{-1}(\{f(x)\}).$$

**Example 4.58.** Suppose $M$ is a subspace of a vector space $V$ and $\sim$ is the equivalence relation of Example 4.54. Then, for $v \in V$,

$$[v] = \{v + z \mid z \in M\}.$$

To prove this, suppose that $w \in [v]$. Then, by definition, $v \sim w$ and so $v - w \in M$. That is, $v - w = z$ for some $z \in M$. Thus $w = v + (-z)$. Therefore $[v] \subseteq \{v + z \mid z \in M\}$.

Now suppose $w = v + z$ for some $z \in M$. Then $v - w = -z = (-1)z \in M$, since $M$ is a subspace. Hence $v \sim w$ and so $w \in [v]$. Hence, $\{v + z \mid z \in M\} \subseteq [v]$.

**Definition 4.59** (Coset)**.** For a subspace $M$ of a vector space $V$ and $v \in V$, we write

$$v + M = \{v + z \mid z \in M\}$$

and call this the *coset of $v$ modulo $M$*.

**Example 4.60.** Define a linear form $f : \mathbb{R}^2 \to \mathbb{R}$ by

$$f(x_1, x_2) = 2x_1 - x_2 \quad \forall\ (x_1, x_2) \in \mathbb{R}^2.$$

Let

$$M = \operatorname{Ker} f = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = 0\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = 0\}.$$

So $M$ is a line through the origin. Note that the range of $f$ is $\mathbb{R}$ since, for instance, for any $x \in \mathbb{R}$, we have $f(0, -x) = x$.

For $c \in \mathbb{R}$, we have

$$f^{-1}(\{c\}) = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = c\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = c\},$$

which is a line parallel to $M$ (and equal to $M$ if and only if $c = 0$).

Consider the equivalence relation of Example 4.53. Under this equivalence relation, for $x, y \in \mathbb{R}^2$, we have

$$x \sim y \iff f(x) = f(y) \iff f(x) - f(y) = 0 \iff f(x - y) = 0 \iff x - y \in \operatorname{Ker} f = M,$$

where we have used the fact that $f$ is linear. Thus we see that this equivalence relation agrees with the one of Example 4.54.

For any $v \in \mathbb{R}^2$,

$$[v] = v + M = f^{-1}(\{f(v)\}) = \{x \in \mathbb{R}^2 \mid 2x_1 - x_2 = f(v)\}.$$

This is the line in $\mathbb{R}^2$ passing through $v$ and parallel to $M$. For example, if $v = (3, 1)$, then $[v]$ is the plane with equation

$$2x_1 - x_2 = 2(3) - (1) = 5.$$

We see that the equivalence relation 'decomposes' the plane into a set of parallel lines. Each point of the plane $\mathbb{R}^2$ lies on exactly one of these lines.

**Definition 4.61** (Partition)**.** Suppose $X$ is a nonempty set. A *partition* of $X$ is a collection $\mathscr{A}$ of subsets of $X$ satisfying the following properties:

(a) every $A \in \mathscr{A}$ is a *nonempty* subset of $X$,
(b) if $A, B \in \mathscr{A}$ and $A \neq B$, then $A \cap B = \varnothing$,
(c) every element of $X$ belongs to one of the subsets in $\mathscr{A}$ (in other words, for each $x \in X$, there is some $A \in \mathscr{A}$ such that $x \in A$).

*Remark* 4.62. The third property of a partition says that each element of $X$ belongs to one of the subsets of the partition. The second property says that no element lies in more than one of the subsets of the partition. Therefore, combining these two facts, we see that the fundamental property of a partition is that every element of $X$ belongs to **exactly one** of the subsets in the partition.

In turns out that equivalence relations and partitions are simply two different ways of thinking about the same thing. The precise relationship is given in the following theorem.

**Theorem 4.63.** *Let $X$ be a nonempty set.*

*(a) If $\sim$ is an equivalence relation on $X$, then the set of equivalence classes is a partition of $X$.*
*(b) Suppose $\mathscr{A}$ is a partition of $X$. Write $x \sim y$ if $x$ and $y$ belong to the same element of $\mathscr{A}$ (that is, there exists an $A \in \mathscr{A}$ such that $x, y \in A$). Then $\sim$ is an equivalence relation on $X$.*

*Proof.* (a) Suppose $\sim$ is an equivalence relation on $X$ and let $\mathscr{A}$ be the set of equivalence classes. We want to show that $\mathscr{A}$ satisfies the conditions of Definition 4.61.

- First of all, for any equivalence class $[x]$, we have $x \in [x]$ (since $x \sim x$ by the reflexivity of an equivalence relation). Therefore, all the equivalence classes are nonempty.
- Suppose $[x]$ and $[y]$ are two equivalence classes and $[x] \neq [y]$. We want to show that $[x] \cap [y] = \varnothing$. We prove this by contradiction. Suppose $[x] \cap [y] \neq \varnothing$. Then we can choose some element $z \in [x] \cap [y]$. We will show that this implies that $[x] = [y]$. Let $w \in [x]$. Then $x \sim w$. Since $z \in [x] \cap [y]$, we have $z \in [x]$ and $z \in [y]$. Thus $x \sim z$ and $y \sim z$. By symmetry, $z \sim x$. Thus

  $$y \sim z \sim x \sim w.$$

  By transitivity, $y \sim w$ and so $w \in [y]$. Hence $[x] \subseteq [y]$. Repeating the above argument but interchanging the roles of $x$ and $y$ shows that $[y] \subseteq [x]$. Hence $[x] = [y]$. This contradicts the assumption that $[x] \neq [y]$. Therefore, $[x] \cap [y] = \varnothing$.
- The last property is easy since for every $x \in X$, we have $x \in [x]$ and so every element of $x$ belongs to some element of $\mathscr{A}$.

(b) Now assume that $\mathscr{A}$ is a partition of $X$ and define $x \sim y$ if there is some $A \in \mathscr{A}$ such that $x, y \in A$. We wish to verify that $\sim$ is an equivalence relation on $X$.
- *Reflexivity*: For any $x \in X$, we have $x \in [x]$. Thus $x \sim x$.
- *Symmetry:* Suppose $x, y \in X$ and $x \sim y$. Thus, there is some $A \in \mathscr{A}$ such that $x, y \in A$. Thus $y \sim x$ as well.
- *Transitivity:* Suppose $x, y, z \in X$, $x \sim y$, and $y \sim z$. Then there are $A, B \in \mathscr{A}$ such that $x, y \in A$ and $y, z \in B$. This implies in particular that $y \in A \cap B$. Thus $A \cap B \neq \varnothing$. Hence, by the definition of a partition, we have $A = B$. Therefore, $x$ and $z$ belong to the same element $A = B$ of $\mathscr{A}$. Hence $x \sim z$.

□

From this theorem, we can deduce some properties of equivalence classes of relations.

**Corollary 4.64.** *Suppose $\sim$ is an equivalence relation on a nonempty set $X$.*

*(a) For any $x, y \in X$, we have either $[x] = [y]$ or $[x] \cap [y] = \varnothing$.*
*(b) For $x, y \in X$, we have $x \sim y$ if and only if $[x] = [y]$.*

*Proof.* (a) This follows immediately from Theorem 4.63, which says that the equivalence classes form a partition.

(b) If $x \sim y$, then $y \in [x]$. We also have $y \in [y]$. Thus $[x] \cap [y] \neq \varnothing$ and so $[x] = [y]$ by the first part of the corollary. Conversely, if $[x] = [y]$, then $y \in [x]$ (since $y \in [y]$). Hence $x \sim y$.

□

We can apply this corollary to the equivalence relation of Example 4.54 to get the following result.

**Corollary 4.65.** *If $M$ is a subspace of a vector space $V$, then*

*(a) for any $x, y \in V$, we have $x + M = y + M$ or $(x + M) \cap (y + M) = \varnothing$.*
*(b) $x - y \in M$ if and only if $x + M = y + M$, and*

**Definition 4.66** (Quotient set)**.** Suppose $\sim$ is an equivalence relation on a set $X$. Then the set of equivalence classes

$$\{[x] \mid x \in X\}$$

is called the *quotient set* of $X$ for the relation $\sim$, and is denoted $X/\sim$. In the special case where $M$ is a subspace of $V$ and the equivalence relation is the one of Example 4.54, the quotient set is denoted $V/M$. Thus

$$V/M = \{x + M \mid x \in V\}.$$

**Example 4.67.** Consider the situation of Example 4.60, where $f : \mathbb{R}^2 \to \mathbb{R}$ is the linear form defined by

$$f(x_1, x_2) = 2x_1 - x_2 \quad \forall\ (x_1, x_2) \in \mathbb{R}^2,$$

and

$$M = \operatorname{Ker} f = \{(x_1, x_2) \in \mathbb{R}^2 \mid f(x_1, x_2) = 0\} = \{(x_1, x_2) \in \mathbb{R}^2 \mid 2x_1 - x_2 = 0\}.$$

Then $\mathbb{R}^2/M$ is the set of lines parallel to $M$.

**Definition 4.68** (Quotient map)**.** Suppose $\sim$ is an equivalence relation on a set $X$. Then we have a surjective map

$$q : X \to X/\sim, \quad q(x) = [x],$$

called the *quotient map* of $X$ onto $X/\sim$.

*Remark* 4.69. Note that if $q : X \to X/\sim$ is the quotient map corresponding to some equivalence relation $\sim$ on a set $X$, then

$$q(x) = q(y) \iff [x] = [y] \iff x \sim y.$$

Thus, any equivalence relation is of the type seen in Example 4.53.

**Lecture 11: October 18, 2010**

4.6. **Quotient vector spaces.** In this course, the most important quotient set will be $V/M$ for a subspace $M$ of a vector space $V$. One reason for this is that, in this case, the quotient set is more than just a **set**: it is a vector space.

First, we need to define vector addition and scalar multiplication. Suppose $V$ is a vector space over a field $F$ and $M$ is a subspace of $V$. For $x + M, y + M \in V/M$, we define their sum to be

$$(x + M) + (y + M) = (x + y) + M.$$

However, this definition should worry you. Why? Note that, in general, an element of $V/M$ can be written in the form $x + M$ for more than one $x$. Since our definition of vector addition refers explicitly to this $x$, our vector addition is not well-defined unless we get the result of our operation is independent of this choice of $x$. That is, if

$$x + M = x' + M \quad \text{and} \quad y + M = y' + M,$$

then it must be true that

$$(x + M) + (y + M) = (x' + M) + (y' + M).$$

By our definition of addition, this is true if and only if

$$(x + y) + M = (x' + y') + M.$$

Let's check this. Since $x + M = x' + M$, we have $x - x' \in M$. Similarly, since $y + M = y' + M$, we have $y - y' \in M$. Thus

$$(x + y) - (x' + y') = (x - x') + (y - y') \in M,$$

since $M$ is a subspace of $V$. Hence $(x + y) + M = (x' + y') + M$ as desired.

We now define scalar multiplication on $V/M$ by the formula

$$c(x + M) = cx + M, \quad c \in F,\ x \in V.$$

Again we need to check that this is well-defined. In particular, we need to check that if $x + M = x' + M$, then $cx + M = cx' + M$. To see this, note that $x + M = x' + M$ implies $x - x' \in M$. Thus $c(x - x') \in M$ since $M$ is a subspace (and hence closed under scalar multiplication). Thus $cx + M = cx' + M$ as desired.

Now that we've defined an addition and scalar multiplication, we can prove the following theorem.

**Theorem 4.70.** *Suppose $V$ is a vector space over a field $F$ and $M$ is a subspace of $V$. Then, under the operations defined above, $V/M$ is a vector space over $F$ and the quotient map*

$$Q : V \to V/M, \quad Q(x) = x + M,$$

*is a linear map with* $\operatorname{Ker} Q = M$. *The vector space $V/M$ is called the* quotient vector space *of $V$ by the subspace $M$.*

*Proof.* We must check that our operations satisfy the axioms of Definition 3.1.

(a) For $x + M, y + M, z + M \in V/M$, we have

$$((x + M) + (y + M)) + (z + M) = ((x + y) + M) + (z + M)$$
$$= (x + y + z) + M = (x + M) + ((y + z) + M) = (x + M) + ((y + M) + (z + M)),$$

so the addition is associative. Also

$$(x + M) + (y + M) = (x + y) + M = (y + x) + M = (y + M) + (x + M), \text{ and}$$
$$(x + M) + (\mathbf{0} + M) = (x + \mathbf{0}) + M = x + M.$$

Thus $(V/M, +)$ is a commutative monoid with identity $\mathbf{0} + M = M$.

(b) For all $x + M \in V/M$, we have

$$(x + M) + (-x + M) = (x - x) + M = \mathbf{0} + M,$$

and so every element of $V/M$ has an additive inverse.

**Exercise:** Check the remaining axioms in the definition of a vector space.

To check that the quotient map $Q$ is linear, suppose $u, v \in V$ and $c \in F$. Then

$$Q(u + v) = (u + v) + M = (u + M) + (v + M) = Q(u) + Q(v),$$
$$Q(cv) = cv + M = c(v + M).$$

□

## 4.7. **The first isomorphism theorem.**

**Lemma 4.71.** *If $T : V \to W$ is a linear mapping, then*

$$T^{-1}(\{Tv\}) = v + \operatorname{Ker} T, \quad \text{for all } v \in V.$$

*Proof.* Suppose $u \in T^{-1}(\{Tv\})$. Then

$$Tu = Tv \implies Tu - Tv = \mathbf{0} \implies T(u - v) = \mathbf{0} \implies u - v \in \operatorname{Ker} T \implies u \in v + \operatorname{Ker} T.$$

Now suppose $u \in v + \operatorname{Ker} T$. Then

$$u - v \in \operatorname{Ker} T \implies T(u - v) = \mathbf{0} \implies Tu - Tv = \{\mathbf{0}\} \implies Tu = Tv \implies u \in T^{-1}(\{Tv\}).$$

□

**Theorem 4.72** (First Isomorphism Theorem)**.** *Suppose $T : V \to W$ is a linear mapping and $N = \operatorname{Ker} T$. Then*

*(a) $N$ is a subspace of $V$,*
*(b) $T(V)$ is a subspace of $W$, and*
*(c) The map $V/N \to T(V)$ given by $v + N \mapsto Tv$ is an isomorphism. Thus, $V/N \cong T(V)$.*

*In other words,*

$$V/\operatorname{Ker} T \cong \operatorname{Im} T$$

*for every linear map $T$ with domain $V$.*

*Proof.* We've already proven (a) and (b) (see Corollary 4.14). So it remains to prove (c).

We need to find a bijective linear mapping $S : V/N \to T(V)$. Suppose $x \in V/N$. Then $x$ is a subset of $V$ and is equal to $v + N$ for some $v \in V$. By Lemma 4.71, $T$ is constant on the subset $x$. Thus we can define a mapping

$$S : V/N \to T(V), \quad S(v + N) = Tv.$$

Note that the map $S$ is only well-defined by the comments above. We first show that $S$ is linear. Suppose $x, y \in V/N$ and $c$ is a scalar. Then $x = u+N$ and $y = v+N$ for some $u, v \in V$. Thus,

$$S(x + y) = S(u + v + N) = T(u + v) = Tu + Tv = S(u + N) + S(v + N) = S(x) + S(y),$$
$$S(cx) = S(cu + N) = T(cu) = cTu = cS(u + N) = cS(x).$$

It remains to show that $S$ is bijective. It is clearly surjective since any element of $T(V)$ is of the form $Tv$ for some $v \in V$ and we have $S(v + N) = Tv$. Since $S$ is linear, we can show it is injective by proving that $\operatorname{Ker} S = \{\mathbf{0}\}$. Now, if $x = u + N \in \operatorname{Ker} S$, then

$$S(u + N) = \mathbf{0} \implies Tu = \mathbf{0} \implies u \in \operatorname{Ker} T \implies u + N = N,$$

which is the zero vector of $V/N$. □

*Remark* 4.73. One of the main uses of the first isomorphism theorem is the following. If we want to prove that $V/U \cong W$ (where $V$ and $W$ are vector spaces and $U$ is a subspace of $V$), then we can do this by finding a surjective linear map from $V$ to $W$ with kernel $U$.

**Example 4.74.** Let

$$U = \{(x_1, x_2, x_3, x_4) \mid 2x_1 - x_2 = 0 \text{ and } x_1 + 3x_2 - 4x_3 - x_4 = 0\}.$$

Prove that $\mathbb{R}^4/U \cong \mathbb{R}^2$.

We need to find a surjective map $T : \mathbb{R}^4 \to \mathbb{R}^2$ such that $U = \operatorname{Ker} T$. Let

$$T(x_1, x_2, x_3, x_4) = (2x_1 - x_2, x_1 + 3x_2 - 4x_3 - x_4).$$

This is a linear map since it corresponds to multiplication by the matrix

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

It is surjective since for any $(a, b) \in \mathbb{R}^2$, we have

$$T\left(\frac{a}{2}, 0, 0, \frac{a}{2} - b\right) = (a, b).$$

(Recalling MAT 1341, showing that $T$ is surjective is equivalent to showing that the above matrix has rank 2). It is clear that $\operatorname{Ker} T = U$. Thus $\mathbb{R}^4/U \cong \mathbb{R}^2$ by the first isomorphism theorem.

**Example 4.75.** Let $X$ be a set and $Y$ be a subset of $X$. Let $F$ be a field and define

$$W = \{f \in \mathcal{F}(X, F) \mid f(x) = 0 \text{ for all } x \in Y\}.$$

Let's prove that $\mathcal{F}(X, F)/W \cong \mathcal{F}(Y, F)$.

First of all, it only makes sense to write $\mathcal{F}(X, F)/W$ if $W$ is a subspace of $\mathcal{F}(X, F)$. But $W$ is just the subset of $\mathcal{F}(X, F)$ consisting of functions equal to zero on the subset $Y$ of $X$. We know this is a subspace by Question 4 of Assignment 2.

Next, we want to find a surjective map $T : \mathcal{F}(X, F) \to \mathcal{F}(Y, F)$ such that $\operatorname{Ker} T = W$. Define $T$ by

$$T(f) = f|_Y.$$

So $T$ restricts a function to $Y$. This map is linear (**exercise**). It is clear that $\operatorname{Ker} T = W$. Also, $T$ is surjective since, given any $F$-valued function $f$ on $Y$, we can extend it to a function $g$ on all of $X$ by given it any values we want on points of $X \setminus Y$. Then $T(g) = f$. Thus, by the first isomorphism theorem, we have $\mathcal{F}(X, F)/W \cong \mathcal{F}(Y, F)$.

**Lecture 12: November 1, 2010**

No office hours on Wednesday, November 3. Instead, extra office hours on Tuesday, November 2, 11:00am–12:00pm.

## 5. Structure of vector spaces

5.1. **Spans and generating sets.** Recall (Definition 3.25) that if $A$ is a nonempty subset of a vector space $V$ over a field $F$, then

$$\operatorname{Span} A = \{c_1v_1 + c_2v_2 + \cdots + c_nv_n \mid n \in \mathbb{N},\ c_i \in F,\ v_i \in A \text{ for } 1 \le i \le n\}.$$

We sometimes write $\operatorname{Span}_F A$ when we want to emphasize the field.

*Remark* 5.1. The text sometimes uses the notation $[A]$ for $\operatorname{Span} A$.

**Theorem 5.2.** *Suppose $A$ is a nonempty subset of a vector space $V$. Then $\operatorname{Span} A$ is a subspace of $V$ and is the smallest subspace of $V$ containing $A$. In other words*

*(a) $A \subseteq \operatorname{Span} A$, and*
*(b) if $W$ is a subspace of $V$ such that $A \subseteq W$, then $\operatorname{Span} A \subseteq W$.*

*Proof.* We already noted in Theorem 3.33 that $\operatorname{Span} A$ is a subspace of $V$, so it remains to proves that it is the smallest one containing $A$. It is clear that $A \subseteq \operatorname{Span} A$ since every element $a \in A$ is a linear combination of elements of $A$ (with one term and coefficient one). Now suppose $W$ is a subspace of $V$ containing $A$. We wish to show that $\operatorname{Span} A \subseteq W$. Let $v \in \operatorname{Span} A$. Then, by definition,

$$v = \sum_{i=1}^{n} c_iv_i$$

for some $c_1, \ldots, c_n \in F$ and $v_1, \ldots, v_n \in A$. Since $A \subseteq W$, each $v_i \in W$, for $1 \le i \le n$. Then, since $W$ is a subspace and therefore closed under scalar multiples and vector addition, $\sum_{i=1}^{n} c_iv_i \in W$. So $v \in W$. Thus $\operatorname{Span} A \subseteq W$ as desired. □

*Remark* 5.3. If we adopt the convention that $\operatorname{Span} \varnothing = \{\mathbf{0}\}$, then Theorem 5.2 remains true with the word 'nonempty' removed.

**Definition 5.4** (Generating set)**.** The space $\operatorname{Span} A$ is called the subspace of $V$ *generated* by $A$, we call $A$ a *generating set* for $\operatorname{Span} A$, and say that $A$ *generates* $\operatorname{Span} A$. So if $\operatorname{Span} A = V$, then we say $A$ generates $V$.

**Theorem 5.5.** *Suppose $T : V \to W$ is a linear map and $A$ is a subset of $V$. Then $\operatorname{Span} T(A) = T(\operatorname{Span} A)$.*

*Proof.* Since $A \subseteq \operatorname{Span} A$, we have $T(A) \subseteq T(\operatorname{Span} A)$. Also, since $\operatorname{Span} A$ is a subspace of $A$, we know $T(\operatorname{Span} A)$ is a subspace of $W$ by Theorem 4.12. Hence, by Theorem 5.2, $\operatorname{Span} T(A) \subseteq T(\operatorname{Span} A)$.

It remains to prove the reverse in inclusion. Since $T(A) \subseteq \operatorname{Span} T(A)$, we have $A \subseteq T^{-1}(\operatorname{Span} T(A))$. Also, since $\operatorname{Span} T(A)$ is a subspace of $W$, we know $T^{-1}(\operatorname{Span} T(A))$ is a subspace of $V$ by Theorem 4.12. Thus, by Theorem 5.2, we have $\operatorname{Span} A \subseteq T^{-1}(\operatorname{Span} T(A))$. Thus $T(\operatorname{Span} A) \subseteq \operatorname{Span} T(A)$. □

**Corollary 5.6.** *If $T : V \to W$ is a surjective linear map and $A$ generates $V$, then $T(A)$ generates $W$.*

*Proof.* We have

$$\begin{aligned} W &= T(V) && \text{(since } T \text{ is surjective)} \\ &= T(\operatorname{Span} A) && \text{(since } A \text{ generates } V) \\ &= \operatorname{Span} T(A) && \text{(by Theorem 5.5).} \end{aligned}$$

Therefore $T(A)$ generates $W$. □

5.2. **Linear dependence/indepdence.**

**Definition 5.7** (Linearly dependent/independent)**.** Suppose $v_1, v_2, \dots, v_n$ is a finite list of vectors in a vector space $V$. We say that the list is (or the vectors $v_1, v_2, \dots, v_n$ themselves are) *linearly dependent* (or simply *dependent*) if there exist scalars $c_1, c_2, \dots, c_n$ such that

$$c_1v_1 + c_2v_2 + \cdots + c_nv_n = \mathbf{0}.$$

and at least one of the $c_i$, $1 \leq i \leq n$, is nonzero. Such an equation (in which at least one of the $c_i$ is nonzero), is called a *dependence relation* among the $v_i$.

If $v_1, v_2, \dots, v_n$ are not linearly dependent, they are said to be *linearly independent*. In other words, the vectors $v_1, v_2, \dots, v_n$ are linearly independent if

$$c_1v_1 + \cdots + c_nv_n = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0,$$

where $c_1, \dots, c_n$ are scalars.

*Remark* 5.8. By the commutativity of vector addition, the order of the vectors in the list is not important in the definition of linear dependence/independence.

**Example 5.9.** Consider the case $n = 1$. The list $v_1$ is linearly dependent if and only if $v_1 = \mathbf{0}$. This is because $c_1v_1 = \mathbf{0}$ for some **nonzero** scalar $c_1$ if and only if $v_1 = 0$ (by Theorem 3.18).

**Example 5.10.** Consider the case $n = 2$. Then the list $v_1, v_2$ is linearly dependent if and only if one of the vectors is a multiple of the other. To see this, first suppose that $v_2 = cv_1$. Then $cv_1 + (-1)v_2 = \mathbf{0}$ is a dependence relation and so $v_1, v_2$ are linearly dependent. The same argument works for the case that $v_1$ is a multiple of $v_2$. Now suppose that $v_1, v_2$ are linearly dependent. Then we have a dependence relation $c_1v_1 + c_2v_2 = \mathbf{0}$ where $c_1$ or $c_2$ (or both) is nonzero. If $c_1 \neq 0$, then $v_1 = (-c_2/c_1)v_2$ and so $v_1$ is a multiple of $v_2$. Similarly, if $c_2 \neq 0$, then $v_2$ is a multiple of $v_1$.

**Lemma 5.11.** *(a) Any list of vectors containing the zero vector is linearly dependent.*
*(b) If the vectors $v_1, v_2, \dots, v_n$ are not distinct (i.e. some vector appears more than once in the list), then they are linearly dependent.*

*Proof.* Consider a list of vectors $v_1, \dots, v_n$.

(a) If $v_i = \mathbf{0}$ for some $i$, $1 \leq i \leq n$, then

$$0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_n$$

is a dependence relation and so the vectors are linearly dependent.

(b) If $v_i = v_j$ for some $1 \leq i < j \leq n$, then

$$0v_1 + \cdots + 0v_{i-1} + 1v_i + 0v_{i+1} + \cdots + 0v_{j-1} + (-1)v_j + 0v_{j+1} + \cdots + v_n$$

is a dependence relation and so the vectors are linearly dependent.

□

**Corollary 5.12.** *Consider a list of vectors $v_1, \dots, v_n$.*

*(a) If $v_1, \dots, v_n$ are linearly independent, then none of them is the zero vector.*
*(b) If $v_1, \dots, v_n$ are linearly independent, then no two of them are equal.*

**Theorem 5.13.** *Suppose $V$ is a vector space over a field $F$ and $v_1, v_2, \dots, v_n \in V$. Define a linear map $T : F^n \to V$ by*

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

*(We know from Theorem 4.4 that this map is linear.) Then the vectors $v_1, \dots, v_n$ are linearly dependent if and only if $T$ is not injective.*

*Proof.* Since $T$ is a linear map, it is not injective if and only if $\operatorname{Ker} T \neq \{\mathbf{0}\}$. This is true if and only if there is some $(a_1, \dots, a_n) \neq (0, \dots, 0)$ such that

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n = \mathbf{0}.$$

This is true if and only if $v_1, \dots, v_n$ are linearly dependent. □

**Corollary 5.14.** *Under the hypotheses of Theorem 5.13, the map $T$ is injective if and only if the vectors $v_1, \dots, v_n$ are linearly independent.*

**Theorem 5.15.** *Let $V$ be a vector space and $v_1, \dots, v_n \in V$, $n \geq 2$. Then the following statements are equivalent:*

*(a) $v_1, \dots, v_n$ are linearly dependent,*
*(b) some $v_j$ is a linear combination of the $v_i$ with $i \neq j$ (i.e. some vector is a linear combination of the others),*
*(c) either $v_1 = \mathbf{0}$ or there exists an index $j > 1$ such that $v_j$ is a linear combination of $v_1, \dots, v_{j-1}$.*

*Proof.* You saw this result for real vector spaces in MAT 1341. The proof for vector spaces over an arbitrary field is essentially the same and so we will omit it. You can also read it in the text (Theorems 3.2.5 and 3.2.7). □

**Corollary 5.16.** *Let $V$ be a vector space and $v_1, \dots, v_n \in V$, $n \geq 2$. Then the following statements are equivalent:*

*(a) $v_1, \dots, v_n$ are linearly independent,*
*(b) no $v_j$ is a linear combination of the $v_i$ with $i \neq j$,*
*(c) $v_1 \neq \mathbf{0}$ and for every $j > 1$, $v_j$ is not a linear combination of the $v_1, \dots, v_{j-1}$.*

**Example 5.17.** The vectors $e_1, \dots, e_n \in F^n$ are linearly independent. This is because

$$c_1 e_1 + \cdots + c_n e_n = \mathbf{0} \implies (c_1, \dots, c_n) = \mathbf{0} \implies c_1 = c_2 = \cdots = c_n = 0.$$

**Lecture 13: November 4, 2010**

**Example 5.18.** Consider the vectors $\sin x, \sin 2x$ in $\mathcal{F}(\mathbb{R})$. Suppose

$$a \sin x + b \sin 2x = 0 \quad \forall\, x \in \mathbb{R}.$$

Note that we insist that the equality hold for **all** $x \in \mathbb{R}$ since equality in the vector space $\mathcal{F}(\mathbb{R})$ is equality **of functions**. So it must hold, in particular, for $x = \pi/2$, and so

$$a \sin \frac{\pi}{2} + b \sin \frac{2\pi}{2} = 0 \implies a + 0 = 0 \implies a = 0.$$

Then, taking $x = \pi/4$, gives

$$a \sin \frac{\pi}{4} + b \sin \frac{2\pi}{4} = 0 \implies 0 \sin \frac{\pi}{4} + b = 0 \implies b = 0.$$

Thus, the vectors $\sin x, \sin 2x$ are linearly independent.

**Theorem 5.19.** *If $T : V \to W$ is an injective linear mapping and $v_1, \dots, v_n$ are linearly independent vectors in $V$, then $Tv_1, \dots, Tv_n$ are linearly independent in $W$.*

*Proof.* For scalars $c_1, \dots, c_n$, we have

$$\begin{aligned} & c_1(Tv_1) + \cdots + c_n(Tv_n) = \mathbf{0} \\ \implies & T(c_1v_1 + \cdots + c_nv_n) = \mathbf{0} && (T \text{ is linear}) \\ \implies & c_1v_1 + \cdots + c_nv_n = \mathbf{0} && (T \text{ is injective}) \\ \implies & c_1 = c_2 = \cdots = c_n = 0 && (\text{the vectors } v_1, \dots, v_n \text{ are linearly independent}). \end{aligned}$$

□

Note that it is very important in the above theorem that $T$ be injective. For instance, it would certainly fail to be true if $T$ were the zero map since then the list $Tv_1, \dots, Tv_n$ would contain the zero vector and so could not be linearly independent.

**Definition 5.20** (Linear dependence/independence of sets)**.** If $A$ is any (possibly infinite) set of vectors in a vector space $V$, we say $A$ is *linearly dependent* if some finite list $v_1, v_2, \dots, v_n$ of distinct vectors in $A$ is linearly dependent (in other words, there is some dependence relation involves a finite number of the vectors of $A$). We say $A$ is *linearly independent* if every finite list of distinct vectors in $A$ is linearly independent.

We will need the following theorem later.

**Theorem 5.21.** *Suppose $V$ is a vector space and $v_1, \dots, v_n \in V$. Let $1 \leq r < n$, let $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, and let $Q : V \to V/M$ be the quotient map. Then the following conditions are equivalent:*

*(a) $v_1, \dots, v_n$ are linearly independent in $V$,*
*(b) $Qv_1, \dots, Qv_r$ are independent in $V/M$ and $v_{r+1}, \dots, v_n$ are linearly independent in $V$.*

*Proof.* (a) $\implies$ (b): $v_{r+1}, \dots, v_n$ are clearly independent since $v_1, \dots, v_n$ are (any subset of an independent set is independent). So it remains to show that $Qv_1, \dots, Qv_n$ are independent. Suppose

$$c_1(Qv_1) + \cdots + c_r(Qv_r) = Q\mathbf{0}$$

(Recall that $Q\mathbf{0} = M$ is the zero vector of $V/M$.) We need to show that $c_1 = \cdots = c_r = 0$. Since $Q$ is linear, we have

$$Q(c_1v_1 + \cdots + c_nv_n) = Q\mathbf{0},$$

and so

$$c_1v_1 + \cdots + c_nv_n \in \operatorname{Ker} Q = M.$$

Since $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, we thus have

$$c_1v_1 + \cdots + c_nv_n = c_{r+1}v_{r+1} + \cdots + c_nv_n,$$

for some scalars $c_{r+1}, \dots, c_n$. Then

$$c_1v_1 + \cdots + c_rv_r + (-c_{r+1})v_{r+1} + \cdots + (-c_n)v_n = \mathbf{0}.$$

Since $v_1, \dots, v_n$ are independent, we have that all the coefficients are zero.

(b) $\implies$ (a): Suppose

$$c_1 v_1 + \cdots + c_n v_n = \mathbf{0}. \tag{5.1}$$

We want to show $c_1 = \cdots = c_n = 0$. Let

$$z = c_{r+1} v_{r+1} + \cdots + c_n v_n.$$

Then $z \in M = \operatorname{Ker} Q$. Therefore,

$$Q\mathbf{0} = c_1(Qv_1) + \cdots + c_r(Qv_r) + Qz = c_1(Qv_1) + \cdots + c_r(Qv_r).$$

Since $Qv_1, \dots, Qv_r$ are independent, we have $c_1 = \cdots = c_r = 0$. But then (5.1) becomes

$$c_{r+1} v_{r+1} + \cdots + c_n v_n = \mathbf{0}.$$

Since $v_{r+1}, \dots, v_n$ are linearly independence, we have $c_{r+1} = \cdots = c_n = 0$. Hence all the $c_i$ are zero. □

## 5.3. **Finitely generated vector spaces.**

**Definition 5.22** (Finitely generated vector space)**.** A vector space $V$ over a field $F$ is *finitely generated* if there is a finite subset $\{v_1, \dots, v_n\} \subseteq V$ such that $\operatorname{Span}\{v_1, \dots, v_n\} = V$. In other words, $V$ is finitely generated if it can be spanned by finitely many vectors. In the case where $V$ can be considered as a vector space over multiple fields (for example, $\mathbb{C}$ is a vector space over both $\mathbb{R}$ and $\mathbb{C}$), we say $V$ is *finitely generated over* $F$ if there is a finite subset $\{v_1, v_2, \dots, v_n\} \subseteq V$ such that $\operatorname{Span}_F\{v_1, \dots, v_n\} = V$.

*Remark* 5.23. Recall (Definition 5.4) that if $\operatorname{Span}\{v_1, \dots, v_n\} = V$, then we say that the set $\{v_1, \dots, v_n\}$ generates $V$. This explains the term 'finitely generated'. Later, once we've defined dimension, we'll often use the term *finite-dimensional* instead of *finitely generated*.

**Examples 5.24.**

(a) The zero space $\{\mathbf{0}\}$ is finitely generated since $\operatorname{Span}\{\mathbf{0}\} = \{\mathbf{0}\}$.
(b) $F^n$ is finitely generated over $F$ since $\operatorname{Span}_F\{e_1, e_2, \dots, e_n\} = F$.
(c) $\mathcal{P}_n(F) = \operatorname{Span}_F\{1, t, t^2, \dots, t^n\}$ and so $\mathcal{P}_n(F)$ is finitely generated over $F$.
(d) $\mathbb{C}$ is finitely generated over $\mathbb{C}$ since $\mathbb{C} = \operatorname{Span}_{\mathbb{C}}\{1\}$.
(e) $\mathbb{C}$ is finitely generated over $\mathbb{R}$ since $\mathbb{C} = \operatorname{Span}_{\mathbb{R}}\{1, i\}$.
(f) $\mathbb{R}$ is finitely generated over $\mathbb{R}$ since $\mathbb{R} = \operatorname{Span}_{\mathbb{R}}\{1\}$.
(g) $\mathbb{R}$ is **not** finitely generated over $\mathbb{Q}$ (but this is not so easy to see).
(h) $\mathcal{F}(\mathbb{R})$ is **not** finitely generated over $\mathbb{R}$.

**Example 5.25.** The vector space $V = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\}$ is finitely generated over $\mathbb{R}$. Indeed, we can show that $V = \operatorname{Span}_{\mathbb{R}}\{\sin, \cos\}$. To do this, we need to prove that any $f \in V$ can be written as a linear combination of sin and cos. Suppose $f \in V$ and let

$$g(x) = f(x)\sin x + f'(x)\cos x.$$

Then

$$g'(x) = f'(x)\sin x + f(x)\cos x + f''(x)\cos x - f'(x)\sin x = (f''(x) + f(x))\cos x = 0.$$

Therefore, $g(x)$ is constant. That is $g(x) = a$ (for all $x \in \mathbb{R}$) for some $a \in \mathbb{R}$. Similarly, if

$$h(x) = f(x)\cos x - f'(x)\sin x,$$

then one can show that $h'(x) = 0$ (**exercise**) and so $h(x) = b$ (for all $x \in \mathbb{R}$) for some $b \in \mathbb{R}$ (that is, $h$ is a constant function). Therefore we have

$$\underbrace{\begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix}}_{A} \begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} = \begin{bmatrix} a \\ b \end{bmatrix}.$$

Since $\det A = -\sin^2 x - \cos^2 x = -1$, for all $x \in \mathbb{R}$, the matrix $A$ is invertible with inverse

$$A^{-1} = \frac{1}{\det A} \begin{bmatrix} -\sin x & -\cos x \\ -\cos x & \sin x \end{bmatrix} = \begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix}.$$

Multiplying both sides of our matrix equation on the left by $A^{-1}$ gives

$$\begin{bmatrix} f(x) \\ f'(x) \end{bmatrix} = \begin{bmatrix} \sin x & \cos x \\ \cos x & -\sin x \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} a \sin x + b \cos x \\ a \cos x - b \sin x \end{bmatrix}.$$

Thus $f(x) = a \sin x + b \cos x$ for all $x \in \mathbb{R}$.

## Lecture 14: November 8, 2010

**Announcement:** Talk aimed at undergraduates, "Introduction to LaTeX and Mathematical Typesetting", Phil Scott, Thursday, November 11, 2:30pm, Room: KED B005.

**Example 5.26.** If $\{v_1, \dots, v_n\}$ generates $V$, and $v_{n+1} \in V$, then $\{v_1, \dots, v_n, v_{n+1}\}$ is dependent. This is because $v_{n+1}$ must be a linear combination of $v_1, \dots, v_n$ and so $\{v_1, \dots, v_{n+1}\}$ is dependent by Theorem 5.15.

In contrapositive form, we have that if $x_1, \dots, x_n$ are independent, then $x_1, \dots, x_{n-1}$ cannot generate $V$.

**Theorem 5.27.** *Suppose $V$ is a vector space over a field $F$ and $v_1, \dots, v_n \in V$. Define a linear map $T : F^n \to V$ by*

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

*Then $T$ is surjective if and only if $v_1, \dots, v_n$ generate $V$.*

*Proof.* By definition, $T$ is surjective if and only if every vector of $V$ can be written as $a_1 v_1 + \cdots + a_n v_n$. This is true if and only if $\operatorname{Span}\{v_1, \dots, v_n\} = V$. □

**Theorem 5.28.** *If $T : V \to W$ is a surjective linear map and $V$ is finitely generated, then $W$ is finitely generated.*

*Proof.* Since $V$ is finitely generated, there is a finite subset $\{v_1, \dots, v_n\} \subseteq V$ such $\{v_1, \dots, v_n\}$ generates $V$. Then, since $T$ is surjective, $T(A) = \{Tv_1, \dots, Tv_n\}$ generates $W$ by Corollary 5.6. Hence $W$ is finitely generated. □

**Example 5.29.** If if $V$ is a finitely generated vector space and $M$ is a subspace of $V$, then $V/M$ is finitely generated. This is because the quotient map $V \to V/M$ is linear and surjective.

**Theorem 5.30.** *A vector space $V$ over $F$ is finitely generated if and only if there exists a positive integer $n$ and a surjective linear map $T : F^n \to V$.*

*Proof.* If $V$ is finitely generated, then $\operatorname{Span}\{v_1, \dots, v_n\} = V$ for some finite subset $\{v_1, \dots, v_n\} \subseteq V$. Then, by Theorem 5.27, there is a surjective linear map $T : F^n \to V$.

Conversely, if there is a positive integer $n$ and a linear surjective map $T : F^n \to V$, then, by Theorem 5.28, $V$ is finitely generated since $F^n$ is. □

We will need the following theorem later.

**Theorem 5.31.** *Suppose $V$ is a vector space and $v_1, \dots, v_n \in V$. Let $1 \leq r < n$, let $M = \operatorname{Span}\{v_{r+1}, \dots, v_n\}$, and let $Q : V \to V/M$ be the quotient map. The following conditions are equivalent.*

*(a) $v_1, \dots, v_n$ generate $V$,*
*(b) $Qv_1, \dots, Qv_r$ generate $V/M$.*

*Proof.* (a) $\implies$ (b): Suppose $u \in V/M$. We want to show that we can write $u$ as a linear combination of $Qv_1, \dots, Qv_n$. Since $Q$ is surjective, there is a $v \in V$ such that $u = Qv$. Since $v_1, \dots, v_n$ generate $V$, we have

$$v = c_1v_1 + \cdots + c_nv_n$$

for some scalars $c_1, \dots, c_n$. We apply the linear map $Q$ to both sides and use the fact that $Qv_{r+1} = \cdots = Qv_n = Q\mathbf{0}$ (since $M = \operatorname{Ker} Q$) to obtain

$$u = Qv = c_1Qv_1 + \cdots + c_rQv_r.$$

(b) $\implies$ (a): Suppose $v \in V$. We want to show that we can write $v$ as a linear combination of $v_1, \dots, v_n$. Since $Qv_1, \dots, Qv_r$ generated $V/M$, we have

$$Qv = c_1(Qv_1) + \cdots + c_r(Qv_r)$$

for some scalars $c_1, \dots, c_r$. Thus

$$v - (c_1v_1 + \cdots + c_rv_r) \in \operatorname{Ker} Q = M,$$

and so

$$v - (c_1v_1 + \cdots + c_rv_r) = c_{r+1}v_{r+1} + \cdots + c_nv_n$$

for some scalars $c_{r+1}, \dots, c_n$. Therefore, $v = c_1v_1 + \cdots + c_nv_n$. □

## 5.4. **Basis and dimension.**

**Definition 5.32** (Basis)**.** A finite list of vectors $v_1, \dots, v_n$ in a vector space $V$ is called a *basis* of $V$ if it is both independent and generating. In other words, every vector $v \in V$ can be written as a linear combination

$$v = a_1v_1 + \cdots + a_nv_n$$

(since $\operatorname{Span}\{v_1, \dots, v_n\} = V$) and the coefficients $a_1, \dots, a_n$ are **unique** by Corollary 5.14. We also speak of the set $\{v_1, \dots, v_n\}$ being a basis. If we wish to emphasize the field, we say $v_1, \dots, v_n$ is a basis of $V$ *over* $F$.

*Remark* 5.33. The plural of 'basis' is 'bases' (pronounced bay-sees).

**Examples 5.34.**
(a) For any field $F$, the set $\{e_1, \dots, e_n\}$ is a basis of $F^n$, called the *canonical basis* or the *natural basis* of $F^n$.
(b) If $v_1, \dots, v_n$ are independent vectors in a vector space $V$, then $\{v_1, \dots, v_n\}$ is a basis of $\operatorname{Span}\{v_1, \dots, v_n\}$.
(c) $\{1\}$ is a basis for $\mathbb{C}$ over $\mathbb{C}$.
(d) $\{1, i\}$ is a basis for $\mathbb{C}$ over $\mathbb{R}$.

(e) $\{1, t, \dots, t^n\}$ is a basis for $\mathcal{P}_n(\mathbb{R})$ over $\mathbb{R}$.

**Example 5.35.** Suppose $v_1, \dots, v_n$ is a basis of $V$. If $v_{n+1}$ is any vector in $V$, then $v_1, \dots, v_{n+1}$ is dependent (since $v_{n+1}$ is a linear combination of the other vectors and so the set is dependent by Theorem 5.15) and so is not a basis. Additionally, the list $v_1, \dots, v_{n-1}$ is not generating (since if it were, $v_n$ would be a linear combination of the vectors in this shorter list and so $v_1, \dots, v_n$ would not be independent).

**Theorem 5.36.** *Suppose $V$ is a vector space over a field $F$ and $v_1, \dots, v_n$ is a finite list of vectors in $V$. Define a linear map $T : F^n \to V$ by*

$$T(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n.$$

*Then the following statements are equivalent:*

*(a) $v_1, \dots, v_n$ is a basis of $V$,*
*(b) $T$ is bijective.*

*Proof.* If $v_1, \dots, v_n$ is a basis if and only if it is both independent and generating. This is true if and only if $T$ is injective and surjective by Theorem 5.27 and Corollary 5.14. □

**Corollary 5.37.** *A vector space $V$ over a field $F$ has a (finite) basis if and only if $V \cong F^n$ for some positive integer $n$.*

*Proof.* If $V \cong \mathbb{F}^n$, then there is a bijective linear map $T : F^n \to V$. Let $e_1, \dots, e_n$ be the canonical basis of $F^n$. Then $Te_1, \dots, Te_n$ is independent (by Theorem 5.19) and generating (by Corollary 5.6) and hence is a basis of $V$.

Conversely, if $V$ has a basis $v_1, \dots, v_n$, then $V \cong F^n$ by Theorem 5.36. □

**Theorem 5.38.** *Every nonzero finitely generated vector space has a basis.*

*Proof.* Let $V \neq \{\mathbf{0}\}$ be a finitely generated vector space. We know that there is some finite set of vectors that spans $V$. Among all such sets, choose one of minimal size, say $\{u_1, \dots, u_k\}$. So $\{u_1, \dots, u_k\}$ generates $V$. We will show that it is also linearly independent (by contradiction). Suppose that this set is dependent. Then, by Theorem 5.15, there exists a $j$, $1 \leq j \leq k$, such that $u_j \in \text{Span}\{u_1, u_2, \dots, \hat{u}_j, \dots, u_k\}$ (where the notation $\hat{u}_j$ means we omit the vector $u_j$). Then $V = \text{Span}\{u_1, \dots, \hat{u}_j, \dots, u_k\}$, which contradicts the minimality of $k$. Therefore, $\{u_1, \dots, u_k\}$ independent and hence is a basis (since it is also generating). □

*Remarks* 5.39. (a) We will see that that **every** vector space (not just finitely generated ones) has a basis.
(b) The above proof shows that we can get a basis by taking a sublist of some generating set. We just keep removing vectors until the vectors are independent.
(c) A vector space can have more than one basis.

Even though a vector space can have more than one basis, any two bases have the same number of vectors, as we will see.

### Lecture 15: November 11, 2010

**Announcement:** Undergraduate Research Opportunity Program (UROP). Deadline for application for Winter semester is December 3, 2010. For more information, go to

www.research.uOttawa.ca/urop.

**Theorem 5.40.** *Suppose $\{v_1, \dots, v_m\}$ and $\{w_1, \dots, w_n\}$ are two sets of vectors in a vector space $V$. If $\{v_1, \dots, v_m\}$ spans $V$ and $\{w_1, \dots, w_n\}$ is linearly independent, then $m \geq n$.*

*Proof.* Since $\{v_1, \dots, v_m\}$ spans $V$, there are scalars $a_{ij} \in F$, $1 \leq i \leq m$, $1 \leq j \leq n$, such that

$$w_j = \sum_{i=1}^{m} a_{ij} v_i, \quad 1 \leq j \leq m.$$

In (block) matrix notation,

$$\begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \underbrace{\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}}_{A}.$$

Suppose $m < n$ (we'll show this is a contradiction). Then we know $\operatorname{rank} A \leq m < n$, and so the columns of $A$ are linearly dependent. Hence there are scalars $c_1, \dots, c_n$, not all zero, such that

$$\sum_{j=1}^{n} c_j (j\text{-th column of A}) = 0.$$

In other words,

$$\sum_{j=1}^{n} c_j a_{ij} = 0 \quad \forall\ i = 1, \dots, m.$$

So we have

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in F^m.$$

But then

$$\begin{bmatrix} w_1 & \cdots & w_n \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} v_1 & \cdots & v_m \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} = 0 \in V.$$

So $c_1 w_1 + \cdots + c_n w_n = 0$. But since the $c_i$ are not all zero, this contradicts the fact that $\{w_1, \dots, w_n\}$ is linearly independent. So our assumption that $m < n$ is false. Hence $m \geq n$. □

*Remark* 5.41. The above theorem tells us that the size of any spanning set is greater than the size of any linearly independent set (of the same vector space). It is proved in the text in a different way.

**Theorem 5.42.** *If $V$ has a basis with $n$ elements, then **every** basis of $V$ has $n$ elements.*

*Proof.* Suppose $B = \{v_1, \dots, v_n\}$ and $C = \{w_1, \dots, w_m\}$ are both bases of $V$. Since $B$ spans $V$ and $C$ is linearly independent, we have $m \leq n$. Since $C$ spans $V$ and $B$ is linearly independent, we also have $n \leq m$. Hence $m = n$. □

**Definition 5.43** (Dimension)**.** Suppose $V$ is a finitely generated vector space. If $V \neq \{\mathbf{0}\}$, then the number of vectors in any basis of $V$ is called the *dimension* of $V$ and is written $\dim V$. If $V = \{\mathbf{0}\}$, we say $\dim V = 0$. If we wish to emphasize the field $F$, we write $\dim_F V$ for the dimension of $V$ over $F$. Finitely generated vector spaces are also called *finite-dimensional*. If $\dim V = n$, we say $V$ is *$n$-dimensional*. If $V$ is not finite-dimensional, we say it is *infinite-dimensional*.

**Examples 5.44.**

(a) $\dim_F F^n = n$.
(b) $\dim_{\mathbb{C}} \mathbb{C} = 1$.
(c) $\dim_{\mathbb{R}} \mathbb{C} = 2$.
(d) $\dim_{\mathbb{R}} \mathcal{P}_n(\mathbb{R}) = n + 1$.
(e) $\dim_{\mathbb{R}} \mathcal{P}_n(\mathbb{C}) = 2(n + 1)$.

**Lemma 5.45.** *Suppose $V$ is a vector space. If $\{v_1, \dots, v_n\}$ spans $V$, then $\dim V \leq n$. If $\{w_1, \dots, w_m\}$ is linearly independent, then $m \leq \dim V$.*

*Proof.* This follows from Theorem 5.40. □

**Lemma 5.46.** *Suppose $\{v_1, \dots, v_n\}$ is independent and $v$ is a vector. Then $\{v, v_1, \dots, v_n\}$ is independent if and only if $v \notin \operatorname{Span}\{v_1, \dots, v_n\}$.*

*Proof.* We prove the contrapositive: That $\{v, v_1, \dots, v_n\}$ is dependent if and only if $v \in \operatorname{Span}\{v_1, \dots, v_n\}$. We already know by Theorem 5.15 that if $v \in \operatorname{Span}\{v_1, \dots, v_n\}$, then $\{v, v_1, \dots, v_n\}$ is dependent. Now suppose $\{v, v_1, \dots, v_n\}$ is dependent. Then there are scalars $a, a_1, \dots, a_n$, not all zero, such that

$$av + a_1v_1 + \cdots + a_nv_n = 0.$$

If $a = 0$, then $a_1v_1 + \cdots + a_nv_n = 0$. Since $\{v_1, \dots, v_n\}$ is independent, we have $a_1 = \cdots = a_n = 0$. This contradicts the fact that $a, a_1, \dots, a_n$ are not all zero. So $a \neq 0$. Then

$$v = -a^{-1}a_1v_1 - \cdots - a^{-1}a_nv_n \in \operatorname{Span}\{v_1, \dots, v_n\}.$$

□

**Theorem 5.47.** *Suppose $V$ is an $n$-dimensional vector space. Then the following statements are equivalent.*

*(a) $\{v_1, \dots, v_n\}$ spans $V$.*
*(b) $\{v_1, \dots, v_n\}$ is linearly independent.*
*(c) $\{v_1, \dots, v_n\}$ is a basis of $V$.*

*Proof.* (a) ⇒ (b): Suppose $\{v_1, \dots, v_n\}$ spans $V$ but is dependent. Then there exists an $i$ such that $V = \operatorname{Span}\{v_1, \dots, \hat{v}_i, \dots, v_n\}$. But then $V$ has a spanning set with $n - 1$ elements. But this contradicts Lemma 5.45. Thus, (b) holds.

(b) ⇒ (a): Suppose $\{v_1, \dots, v_n\}$ is independent but does not span $V$. Then we can find a vector $v \in V$ such that $v \notin \operatorname{Span}\{v_1, \dots, v_n\}$. Then, by Lemma 5.46, $\{v, v_1, \dots, v_n\}$ is an independent set with $n + 1$ elements. But this contradicts Lemma 5.45. Thus, (a) holds.

It is now clear that (a) and (b) are equivalent to (c). □

*Remark* 5.48. The point of the above theorem is that if we know the dimension of a vector space ahead of time, to show that some set is a basis we only need to check **one** of the two defining properties of a basis (independence or spanning).

**Example 5.49.** We know $\dim_{\mathbb{R}} \mathbb{C} = 2$. Since $\{1, 1+i\}$ are independent (**exercise**), it is a basis.

**Example 5.50.** Let

$$W = \{f \in C^\infty(\mathbb{R}) \mid f'' + f = 0\} = \text{Span}\{\sin, \cos\}.$$

**Exercise:** Show $\sin, \cos$ are linearly independent. Then $\dim W = 2$.
Define

$$f(x) = \sin x + \cos x$$
$$g(x) = \sin x - \cos x.$$

Then $f, g$ are linearly independent (**exercise**). Therefore $\{f, g\}$ is a basis for $W$.

**Example 5.51.** Let

$$U = \{(x, y, z) \in \mathbb{C}^3 \mid x + y + z = 0\} = \text{Span}\{(-1, 1, 0), (-1, 0, 1)\}.$$

(You learned how to solve homogeneous systems like this in MAT 1341.) Moreover, $\{(-1, 1, 0), (-1, 0, 1)\}$ is linearly independent (it is a two vector set, and neither vector is a multiple of the other). Thus $\{(-1, 1, 0), (-1, 0, 1)\}$ is a basis of $U$ and so $\dim U = 2$.

Since $(1, 1, -2)$, $(0, 1, -1)$ both belong to $U$ and are linearly independent, $\{(1, 1, -2), (0, 1, -2)$ is another basis for $U$.

**Theorem 5.52.** *Suppose $V$ and $W$ are finite-dimensional vector spaces. Then $V \cong W$ if and only if $\dim V = \dim W$.*

*Proof.* Suppose $\dim V = n$. Then $V$ has a basis $\{v_1, \ldots, v_n\}$ with $n$ vectors. If $V \cong W$, then there exists an isomorphism $T : V \to W$. Then $Tv_1, \ldots, Tv_n$ generates $W$ (Corollary 5.6) and is independent (Theorem 5.19). Then $\{Tv_1, \ldots, Tv_n\}$ is a basis of $W$ and so $\dim W = n = \dim V$.

Suppose suppose $\dim W = \dim V = n$. Then $W$ has a basis $\{w_1, \ldots, w_n\}$ with $n$ elements. Then, by Theorem 5.36, $V \cong F^n$ and $W \cong F^n$. Hence $V \cong W$ (since isomorphism is an equivalence relation). □

**Theorem 5.53.** *Suppose $W$ is a subspace of a vector space $V$. Then*

*(a) $W$ is finite-dimensional and $\dim W \leq \dim V$, and*
*(b) $\dim W = \dim V$ if and only if $W = V$.*

*Proof.* Let $n = \dim V$. If $W = \{\mathbf{0}\}$, then the theorem is trivially true. Therefore, assume $W \neq \{\mathbf{0}\}$. Hence $V \neq \{\mathbf{0}\}$ and $n \geq 1$. Choose $w_1 \in W$, $w_1 \neq 0$. Then $\{w_1\}$ is independent. If $\text{Span}\{w_1\} = W$, then $W$ is finitely-generated. Otherwise, choose $w_2 \notin \text{Span}\{w_1\}$. Then $\{w_1, w_2\}$ is independent. We continue in the manner, extending the list. By Lemma 5.45, this process must stop and so we obtain a list $\{w_1, \ldots, w_k\}$ which is both independent and spans $W$, with $k \leq n$. Therefore, $\dim W \leq \dim V$. If $\dim W = \dim V$ (i.e. $k = n$), then $\{w_1, \ldots, w_k\}$ must span $V$ since it is independent (Theorem 5.47) and so $W = \text{Span}\{w_1, \ldots, w_k\} = V$. It is clear that if $W = V$, then $\dim W = \dim V$. □

**Corollary 5.54.** *Every linearly independent set of vectors in a finitely-generated vector space $V$ can be extended to a basis of $V$.*

*Proof.* This follows from the method of the proof of the above theorem. □

**Lecture 16: November 15, 2010**

**Example 5.55. (For DGD)** Take

$$V = \mathcal{P}_2(\mathbb{R}) = \{a_0 + a_1 t + a_2 t^2 \mid a_i \in \mathbb{R}\}.$$

The set $\{1-t\}$ is independent in $V$ (since $1-t$ is not the zero polynomial, e.g. $1-t \neq 0$ at $t=0$). Since $\operatorname{Span}\{1-t\} \neq \mathcal{P}_2(\mathbb{R})$, we continue. Choose a polynomial, for instance, $t \in \mathcal{P}_2(\mathbb{R})$, but $t \notin \operatorname{Span}\{1-t\}$. Then $\{1-t, t\}$ is linearly independent. Is $\operatorname{Span}\{1-t, t\} = \mathcal{P}_2(\mathbb{R})$? No (for instance $t^2 \notin \operatorname{Span}\{1-t, t\}$). Then the set $\{1-t, t, t^2\}$ is independent in $\mathcal{P}_2(\mathbb{R})$. Since $\dim \mathcal{P}_2(\mathbb{R}) = 3$, we know that $\{1-t, t, t^2\}$ is a basis.

## 5.5. Conservation of dimension.

**Theorem 5.56.** *If $V$ is a vector space and $M$ is a subspace of $V$, then the following conditions are equivalent:*

*(a) $V$ is finite-dimensional,*
*(b) $M$ and $V/M$ are finite-dimensional.*

*When these conditions hold, we have*

$$\dim V = \dim(V/M) + \dim M. \tag{5.2}$$

*Proof.* (a) $\Rightarrow$ (b): Suppose $V$ is finite-dimensional. Then $M$ is finite-dimensional by Theorem 5.53 and $V/M$ is finite-dimensional by applying Theorem 5.28 to the quotient map $Q : V \to V/M$.

(b) $\Rightarrow$ (a): If $M = V$, the implication is trivial. Furthermore, $V/M = V/V$ consists of the single coset $\mathbf{0} + V = V$ and so $V/V$ is the zero vector space. Thus (5.2) becomes

$$\dim V = 0 + \dim V,$$

which is obviously true.

If $M = \{\mathbf{0}\}$, then the quotient mapping $V \to V/M$ is a linear surjection with zero kernel. Thus it is an isomorphism and so $V \cong V/M$. Thus $V$ is finite-dimensional because $V/M$ is. Furthermore, (5.2) becomes

$$\dim V = \dim(V/\{\mathbf{0}\}) + 0 = \dim V + 0,$$

which is clearly true.

Now assume $M \neq \{\mathbf{0}\}$ and $M \neq V$. Choose a basis $x_1, \dots, x_m$ of $M$ and $u_1, \dots, u_r$ of $V/M$ (we can do this since, by assumption, $M$ and $V/M$ are finite-dimensional). Choose $y_k \in V$ such that $u_k = y_k + M$, $1 \leq k \leq r$. Then, by Theorem 5.31, the list

$$x_1, \dots, x_m, y_1, \dots, y_r$$

generates $V$. Furthermore, this list is independent by Theorem 5.21. Therefore, it is a basis for $V$. It follows that $V$ is finite-dimensional and

$$\dim V = m + r = \dim M + \dim V/M.$$

□

We can generalize the above theorem as follows.

**Theorem 5.57** (Conservation of dimension)**.** *If $V$ is a finite-dimensional vector space and $T : V \to W$ is a linear mapping, then*

$$\dim V = \dim T(V) + \dim(\operatorname{Ker} T).$$

*Proof.* Since $V$ is finite-dimensional, so is its image $T(V)$ ($T$ is a surjection onto its image, and we call apply Theorem 5.28) and the subspace $\operatorname{Ker} T$ (by Theorem 5.53). By the First Isomorphism Theorem, $T(V) \cong V/\operatorname{Ker} T$. Thus

$$\dim T(V) = \dim(V/\operatorname{Ker} T).$$

Now, by Theorem 5.56, we have $\dim T(V) = \dim V - \dim(\operatorname{Ker} T)$. The result follows. □

*Remark* 5.58. Note that Theorem 5.57 does **not** say that $V \cong T(V) \oplus \operatorname{Ker} T$. Most of the time, such an isomorphism doesn't even make sense since $T(V)$ is a subspace of $W$, not $V$.

**Definition 5.59** (Rank and Nullity)**.** If $T : V \to W$ is a linear map, we define the *rank* of $T$ to be

$$\operatorname{rank} T = \dim T(V),$$

and the *nullity* of $T$ to be

$$\operatorname{null} T = \dim(\operatorname{Ker} T).$$

We can now rephrase Theorem 5.57 as

$$\operatorname{rank} T + \operatorname{null} T = \dim V,$$

where $V$ is the domain of the linear map $T$.

**Corollary 5.60.** *If $T : V \to W$ is linear, and $\dim V = \dim W < \infty$, then*

$$T \text{ is an isomorphism} \iff T \text{ is injective} \iff T \text{ is surjective.}$$

*Proof.* Obviously, if $T$ is an isomorphism, it is injective (by definition). Suppose $T$ is injective. Then $\operatorname{Ker} T = \{\mathbf{0}\}$ and so $\dim T(V) = \dim V$ by the conservation of dimension theorem. So $T(V) = V$ (since $T(V)$ is a subspace of $V$, we can apply Theorem 5.53). Finally, suppose $T$ is surjective. Then $T(V) = V$ and so $\operatorname{rank} T = \dim V$. By the conservation of dimension theorem, $\operatorname{null} T = 0$ and so $T$ is injective, hence an isomorphism. □

**Theorem 5.61.** *If $V_1, V_2, \ldots, V_n$ are finite-dimensional vector spaces, then so is $V = V_1 \times \cdots \times V_n$ and*

$$\dim V = \dim V_1 + \dim V_2 + \cdots + \dim V_3.$$

*Proof.* See text (Theorem 3.6.4 and Corollary 3.6.5). □

**Theorem 5.62.** *A system of $m$ homogeneous linear equations in $n$ unknowns, where $n > m$, always has a nontrivial solution.*

*Proof.* As you saw in MAT 1341, such a system is equivalent to a matrix equation

$$Ax = \mathbf{0},$$

where $A$ is the coefficient matrix, and hence is $m \times n$. The map

$$T : F^n \to F^m, \quad T(x) = Ax,$$

is linear (since it is multiplication by a matrix). The set of solutions is precisely the kernel of $T$. By Theorem 5.57, we have

$$\dim T(F^n) + \dim \operatorname{Ker} T = \dim F^n = n \implies \dim \operatorname{Ker} T = n - \dim T(F^n) \geq n - m > 0.$$

Here we used that $T(F^n) \subseteq F^m$ and so $\dim T(F^n) \leq \dim F^m = m$. So $\dim \operatorname{Ker} T > 0$, which means that $\operatorname{Ker} T$ is not the zero vector space and hence there are nonzero elements in the kernel of $T$ (which correspond to nontrivial solutions to the homogeneous system). □

**Theorem 5.63.** *Suppose $T : V \to W$ is a linear mapping between finite-dimensional vector spaces.*

*(a) If $T$ is injective, then $\dim V \leq \dim W$.*
*(b) If $T$ is surjective, then $\dim V \geq \dim W$.*
*(c) If $T$ is bijective, then $\dim V = \dim W$.*

*Proof.* (a) If $T$ is injective, then $\operatorname{Ker} T = \{\mathbf{0}\}$. Thus

$$\dim V = \dim T(V) + \dim \operatorname{Ker} T = \dim T(V) + 0 \leq \dim W.$$

(b) If $T$ is surjective, then $T(V) = W$. Thus

$$\dim V = \dim T(V) + \dim \operatorname{Ker} T = \dim W + \dim \operatorname{Ker} T \geq \dim W.$$

(c) This follows immediately from the previous two parts.

□

## 5.6. **Dimensions of spaces of linear mappings.**

**Theorem 5.64.** *Suppose $V$ is an $n$-dimensional vector space and $W$ is an arbitrary (possibly infinite-dimensional) vector space. If $\{v_1, \dots, v_n\}$ is a basis of $V$ and $w_1, \dots, w_n$ are **any** vectors in $W$, then there exists a unique linear mapping $T : V \to W$ such that $Tv_i = w_i$ for all $i = 1, \dots, n$.*

*Proof. Existence:* We know (Theorem 4.4) that the maps $R : F^n \to V$ and $S : F^n \to W$ defined by

$$R(a_1, \dots, a_n) = a_1 v_1 + \cdots + a_n v_n,$$
$$S(a_1, \dots, a_n) = a_1 w_1 + \cdots + a_n w_n,$$

are linear. It is clear that if $\{e_1, \dots, e_n\}$ is the canonical basis of $F^n$, then

$$Re_i = v_i, \quad Se_i = w_i.$$

Since $\{v_1, \dots, v_n\}$ is a basis of $V$, $R$ is bijective by Theorem 5.36. So $R$ is invertible and $R^{-1}$ is linear by Theorem 4.47. Therefore, the map $T \stackrel{\text{def}}{=} SR^{-1}$ is linear. Since

$$Tv_i = SR^{-1}v_i = Se_i = w_i,$$

the map $T$ has the desired properties.

*Uniqueness:* Suppose $T_1, T_2 : V \to W$ are two linear maps with the given property. Then

$$(T_1 - T_2)(v_i) = T_1 v_i - T_2 v_i = w_i - w_i = \mathbf{0} \quad \forall\ i = 1, \dots, n.$$

Since $\{v_1, \dots, v_n\}$ is a basis for $V$, this means that $T_1 - T_2 = 0$ (the zero map). Thus $T_1 = T_2$. □

**Corollary 5.65.** *If $V$ and $W$ are finite-dimensional vector spaces, then $\mathcal{L}(V, W)$ is also finite-dimensional and*

$$\dim \mathcal{L}(V, W) = (\dim V)(\dim W).$$

*Proof.* Let $n = \dim V$ and let $\{v_1, \dots, v_n\}$ be a basis of $W$. Define a map

$$\Phi : \mathcal{L}(V, W) \to W^n = W \times W \times \cdots \times W, \quad \Phi(T) = (Tv_1, \dots, Tv_n).$$

Then $\Phi$ is linear (**exercise**) and bijective by Theorem 5.64. Hence $\Phi$ is a vector space isomorphism and so

$$\dim \mathcal{L}(V, W) = \dim W^n = n(\dim W) = (\dim V)(\dim W).$$

(In the first equality, we used Theorem 5.61). □

5.7. **Dual spaces.** Recall (Definition 4.10) that if $V$ is a vector space over a field $F$ then *dual space* of linear forms is

$$V^* = \mathcal{L}(V, F).$$

*Remark* 5.66. The text uses the notation $V'$ instead of $V^*$.

**Theorem 5.67.** *If* $\dim V < \infty$, *then* $\dim V^* = \dim V$.

*Proof.* This follows from Corollary 5.65:

$$\dim V^* = \dim \mathcal{L}(V, F) = (\dim V)(\dim F) = (\dim V) \cdot 1 = \dim V.$$

□

**Proposition 5.68** (Existence of dual bases)**.** *Suppose* $\{v_1, \dots, v_n\}$ *is a basis of a vector space* $V$. *Then there exists a basis* $\{f_1, \dots, f_n\}$ *of* $V^*$ *such that*

$$f_i(v_j) = \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases}$$

*(The symbol* $\delta_{ij}$ *is called the* Kronecker delta.*)*

*Proof.* For $i = 1, \dots, n$, define

$$f_i : V \to F, \quad f_i \left( \sum_{j=1}^{n} c_j v_j \right) = c_i.$$

Then each $f_i$ is linear (**exercise**) and so $f_i \in V^*$. Moreover, we easily see that $f_i(v_j) = \delta_{ij}$. Since we know from Theorem 5.67 that $\dim V^* = n$, to show that $\{f_1, \dots, f_n\}$ is a basis, it is enough to show that this set is linearly independent (Theorem 5.47). Now, for $c_1, \dots, c_n \in F$,

$$\begin{aligned} \sum_{j=1}^{n} c_j f_j = 0 &\implies \sum_{j=1}^{n} c_j f_j(v_i) = 0 \ \forall \ i = 1, \dots, n \\ &\implies \sum_{j=1}^{n} c_j \delta_{ji} = 0 \ \forall \ i = 1, \dots, n \\ &\implies c_i = 0 \ \forall \ i = 1, \dots, n. \end{aligned}$$

Thus $\{f_1, \dots, f_n\}$ is linearly independent and so we're done. □

**Lecture 17: November 18, 2010**

**Definition 5.69** (Dual basis)**.** We call $\{f_1, \dots, f_n\}$ the *basis of* $V^*$ *dual to* $\{v_1, \dots, v_n\}$.

**Example 5.70.** If $\{e_1, \dots, e_n\}$ is the canonical basis of $F^n$, then the dual basis of $(F^n)^*$ is $\{f_1, \dots, f_n\}$ where

$$f_i(a_1, \dots, a_n) = a_i$$

is the $i$-th *coordinate function*.

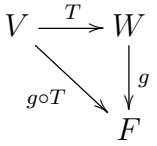
If $f : \mathbb{R}^3 \to \mathbb{R}$ is defined by $f(x, y, z) = x - 2y + 3z$ (so $f \in (\mathbb{R}^3)^*$), then

$$f = f_1 - 2f_2 + 3f_3.$$

**Theorem 5.71.** *Suppose $T : V \to W$ is a linear map. Then the map $T^* : W^* \to V^*$ defined by*

$$T^*g = g \circ T, \quad g \in W^*$$

*is linear.*




*Proof.* First, it is clear that for $g \in W^* = \mathcal{L}(W, F)$, the composition $g \circ T$ is a map from $V$ to $F$. So we only need to show it is linear. Suppose $g, h \in W^*$ and $c, d$ are scalars. Then for all $v \in V$, we have

$$\begin{aligned}
(T^*(cg + dh))(v) &= ((cg + dh)T)(v) \\
&= (cg + dh)(Tv) \\
&= c(g(Tv)) + d(h(Tv)) \\
&= c(gT)(v) + d(hT)(v) \\
&= c(T^*g)(v) + d(T^*h)(v) \\
&= (cT^*g + dT^*h)(v).
\end{aligned}$$

Thus $T^*(cg + dh) = cT^*g + dT^*h$. Therefore $T^*$ is linear. □

**Definition 5.72** (Transpose)**.** The map $T^*$ is called the *transpose* of the map $T$.

*Remark* 5.73. We will see later than the transpose of a linear map is closely related to the transpose of matrix (which you learned about in MAT 1341).

**Theorem 5.74.** *If $T : V \to W$ is a linear map, then*

$$\operatorname{Ker} T^* = \{g \in W^* \mid g(w) = 0 \ \forall \ w \in T(V)\}.$$

*Proof.* If $g \in W^*$, then

$$\begin{aligned}
g \in \operatorname{Ker} T^* &\iff T^*g = 0 \iff gT = 0 \iff (gT)(v) = 0 \ \forall \ v \in V \\
&\iff g(Tv) = 0 \ \forall \ v \in V \iff g(w) = 0 \ \forall \ w \in T(V).
\end{aligned}$$

□

**Definition 5.75** (Annihilator)**.** If $M$ is a subspace of $V$, then the *annihilator* of $M$ in $V^*$ is

$$M^\circ \stackrel{\text{def}}{=} \{f \in V^* \mid f(w) = 0 \ \forall \ w \in M\} = \{f \in V^* \mid f = 0 \text{ on } M\}.$$

Note that $M^\circ$ is a subspace of $V^*$.

Using the above terminology, the result of Theorem 5.74 is

$$\operatorname{Ker} T^* = T(V)^\circ = (\operatorname{Im} T)^\circ.$$

**Theorem 5.76.** *If $T : V \to W$ is a linear map, then $\operatorname{Im} T^* = (\operatorname{Ker} T)^\circ$.*

*Proof.* Let $f \in \operatorname{Im} T^*$. Then $f = T^* g$ for some $g \in W^*$. Thus

$$f(v) = (T^* g)(v) = gTv \ \forall\ v \in V.$$

So

$$v \in \operatorname{Ker} T \implies Tv = 0 \implies f(v) = g(0) = 0.$$

Therefore $\operatorname{Im} T^* \subseteq (\ker T)^\circ$.

Now suppose $f \in (\operatorname{Ker} T)^\circ$. We want to find a $g \in W^*$ such that $f = T^* g$. Let $\{v_1, \dots, v_k\}$ be a basis of $\operatorname{Ker} T$ and extend this to a basis $\{v_1, \dots, v_k, u_1, \dots, u_l\}$ of $V$. We know that $\{Tu_1, \dots, Tu_l\}$ is a basis of $\operatorname{Im} T$. Extend this to a basis of $\{Tu_1, \dots, Tu_l, w_1, \dots, w_p\}$ of $W$. Now define $g \in W^*$ as follows. For scalars $c_1, \dots, c_l, a_1, \dots, a_p$, define

$$g(cTu_1 + \cdots + c_l Tu_l + a_1 w_1 + \cdots + a_p w_p) = c_1 f(u_1) + \cdots + c_l f(u_l) = f(c_1 u_1 + \cdots + c_l u_l).$$

Then $g$ is linear by Theorem 5.64. We will show that $T^* g = f$. Let $x \in V$ and write $x = v + u$ with $v \in \operatorname{Ker} T$ and $u \in \operatorname{Span}\{u_1, \dots, u_l\}$. Then

$$(T^* g)(x) = gT(v + u) = gT(u) = f(u).$$

On the other hand,

$$f(x) = f(v + u) = f(v) + f(u) = f(u),$$

since $f \in (\operatorname{Ker} T)^\circ$. Hence $T^* g = f$. Thus $(\operatorname{Ker} T)^\circ \subseteq \operatorname{Im} T^*$ and so $(\operatorname{Ker} T)^\circ = \operatorname{Im} T^*$. □

**Theorem 5.77.** *If $U$ is a subspace of $V$ and $\dim V < \infty$, then*

$$\dim U^\circ = \dim V - \dim U = \dim V^* - \dim U^*.$$

*Proof.* Let $j : U \to V$ be the inclusion map (i.e. $j(u) = u$ for all $u \in U$). Then we have

$$\operatorname{Ker} j^* = (\operatorname{Im} j)^\circ \quad \text{and} \quad \operatorname{Im} j^* = (\operatorname{Ker} j)^\circ.$$

Now, $\operatorname{Im} j = U$ and $\operatorname{Ker} j = \{\mathbf{0}\}$. Thus

$$\operatorname{Ker} j^* = U^\circ \quad \text{and} \quad \operatorname{Im} j^* = \{\mathbf{0}\}^\circ = U^*$$

(so $j^*$ is surjective). By the conservation of dimension (Theorem 5.57), we have

$$\dim \operatorname{Ker} j^* + \dim \operatorname{Im} j^* = \dim V^*.$$

Thus

$$\dim U^\circ + \dim U^* = \dim V^* = \dim V.$$

Since $\dim U = \dim U^*$, the result follows. □

**Corollary 5.78.** *If $T : V \to W$ is linear, and $\dim V, \dim W < \infty$, then*

*(a)* $\operatorname{rank} T = \operatorname{rank} T^*$,
*(b) $T$ is surjective $\iff$ $T^*$ is injective, and*
*(c) $T$ is injective $\iff$ $T^*$ is surjective.*

*Proof.* (a) We have

$$\begin{aligned}\operatorname{rank} T^* &= \dim W^* - \dim(\operatorname{Ker} T^*) \\ &= \dim W^* - \dim(\operatorname{Im} T)^\circ \\ &= \dim W^* - \big(\dim W - \dim(\operatorname{Im} T)\big) \\ &= \dim \operatorname{Im} T \\ &= \operatorname{rank} T.\end{aligned}$$

(b) We have

$$T \text{ is surjective} \iff \operatorname{rank} T = \dim W \iff \operatorname{rank} T^* = \dim W^*$$
$$\iff \dim \operatorname{Ker} T^* = 0 \iff T^* \text{ is injective.}$$

(c) We have

$$T \text{ is injective} \iff \operatorname{rank} T = \dim V \iff \operatorname{rank} T^* = \dim V^* \iff T^* \text{ is surjective.}$$

□

**Lecture 18: November 22, 2010**

## 6. Matrices

6.1. **The matrix of a linear map.** Much of the material in Section 4 of the text is review from MAT 1341. The only real new part is the matrix of a linear map, which generalizes the "standard matrix" of a map $\mathbb{R}^n \to \mathbb{R}^m$ seen in MAT 1341. Therefore, we will move through Section 4 quite quickly. Students should read through the material we skip as a review.

**Definition 6.1** (The matrix of a linear map)**.** Suppose $V$ and $W$ are finite-dimensional vector spaces over a field $F$, $n = \dim V$, $m = \dim W$, and $T : V \to W$ is a linear map. Let $B = \{v_1, \ldots, v_n\}$ be an ordered basis of $V$ and let $D = \{w_1, \ldots, w_m\}$ be an ordered basis of $W$. For each $j = 1, \ldots, n$, write $Tv_j$ as a linear combination of $y_1, \ldots, y_m$:

$$Tv_j = \sum_{i=1}^{m} a_{ij} w_i, \quad j = 1, \ldots, n.$$

Then the $m \times n$ matrix $[a_{ij}]$ is called the *matrix of $T$ relative to the bases $v_1, \ldots, v_n$ and $w_1, \ldots, w_m$* and is denoted $M_T^{DB}$.

*Remark* 6.2.
(a) It is important that the bases be **ordered**. Changing the order of vectors in the bases will change the matrix.
(b) The text uses the notation $M_T$ instead of $M_T^{DB}$. We will sometimes use this notation when we've fixed the bases $B$ and $D$ and there is no chance of confusion.
(c) A given linear map can have different matrices (if you pick different bases).

Recall that if $V$ is an $n$-dimensional space, and $B = \{v_1, \ldots, v_n\}$ is an ordered basis of $V$, then we have an isomorphism

$$C_B : V \to F^n, \quad C_B\left(\sum_{i=1}^{n} c_i v_i\right) = \begin{bmatrix} c_1 \\ \vdots \\ c_n \end{bmatrix} \in F^n.$$

So $C_B$ is the *coordinate* function that takes the coordinates of a vector relative to the basis $B$. If $D$ is an ordered basis of $W$ and $T \in \mathcal{L}(V, W)$, we have the following diagram.

$$\begin{array}{ccc} V & \xrightarrow{\quad T \quad} & W \\ {\scriptstyle C_B}\downarrow\uparrow{\scriptstyle C_B^{-1}} & & {\scriptstyle C_D}\downarrow\uparrow{\scriptstyle C_D^{-1}} \\ F^n & \xrightarrow[C_D T C_B^{-1}]{} & F^m \end{array}$$

We know from MAT 1341 that a linear map $F^n \to F^m$ corresponds to multiplication by a matrix (in MAT 1341, $F$ was $\mathbb{R}$ or $\mathbb{C}$, but the argument is the same in general). So $C_D T C_B^{-1}$ corresponds to multiplication by some matrix, and this is the matrix $M_T^{DB}$. This gives us an isomorphism

$$\mathcal{L}(V, W) \to \mathcal{L}(F^n, F^m), \quad T \mapsto C_D T C_B^{-1},$$

and an isomorphism

$$\mathcal{L}(V, W) \to M_{mn}(F), \quad T \mapsto M_T^{DB}.$$

We often simply identify an $m \times n$ matrix with the corresponding map $F^m \to F^n$ (given by multiplication by that matrix) and so we write $M_T^{DB} = C_D T C_B^{-1}$.

Note that

$$C_D(Tv_j) = \begin{bmatrix} a_{ij} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

and so

$$M_T^{DB} = \begin{bmatrix} C_D(Tv_1) & C_D(Tv_2) & \cdots & C_D(Tv_n) \end{bmatrix}.$$

**Example 6.3.** Let $S : F^n \to F^m$ be a linear map and choose the standard bases $B = \{e_1, \dots, e_n\}$ and $D = \{e_1, \dots, e_m\}$. Then $C_B : F^n \to F^n$ and $C_D : F^n \to F^n$ are the identity maps. Thus

$$M_T^{DB} = \begin{bmatrix} Se_1 & Se_2 & \cdots & Se_n \end{bmatrix}$$

is the standard matrix of the linear transformation (as you saw in MAT 1341).

Recall (from MAT 1341) that, for a matrix $A$, $\operatorname{col} A$ denotes the *column space* of $A$ (the span of the columns of $A$).

**Proposition 6.4.** *With the same notation as above, we have*

*(a)* $M_T C_B(v) = C_D T(v)$ *for all* $v \in V$,
*(b)* $C_B(\operatorname{Ker} T) = \operatorname{Ker} M_T$ *or* $\operatorname{Ker} T = C_B^{-1}(\operatorname{Ker} M_T)$, *and*
*(c)* $C_D(\operatorname{Im} T) = \operatorname{Im} M_T = \operatorname{col} M_T$ *or* $\operatorname{Im} T = C_D^{-1}(\operatorname{col} M_T)$.

*Hence* $\operatorname{rank} T = \operatorname{rank} M_T$.

*Proof.* These statements all follow from the equation $M_T = C_D T C_B^{-1}$ and the fact that $C_B$ and $C_D$ are isomorphisms. □

This proposition tells us that to compute the rank of any linear transformation, you can compute the rank of its matrix in **any** basis. So we reduce the problem to matrix calculations you learned in MAT 1341.

**Example 6.5.** Let $T : \mathcal{P}_3(\mathbb{R}) \to \mathcal{P}_2(\mathbb{R})$ be the linear map given by $T(p) = p' - p''$. Choose ordered bases $B = \{1, t, t^2, t^3\}$ and $D = \{1, t, t^2\}$ of $\mathcal{P}_3(\mathbb{R})$ and $\mathcal{P}_2(\mathbb{R})$ respectively. Then

$$\begin{aligned} M_T &= \begin{bmatrix} C_D T(1) & C_D T(t) & C_D T(t^2) & C_D T(t^3) \end{bmatrix} \\ &= \begin{bmatrix} C_D(0) & C_D(1) & C_D(2t-2) & C_D(3t^2-6t) \end{bmatrix} \\ &= \begin{bmatrix} 0 & 1 & -2 & 0 \\ 0 & 0 & 2 & -6 \\ 0 & 0 & 0 & 3 \end{bmatrix}. \end{aligned}$$

Now let's find the kernel and image of $T$. Using techniques from MAT 1341, we know that

$$\operatorname{Ker} M_T = \operatorname{Span}\left( \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} \right), \quad \operatorname{col} M_T = \mathbb{R}^3.$$

Therefore

$$\operatorname{Ker} T = \operatorname{Span}\left\{ C_B^{-1}\left(\begin{bmatrix}1\\0\\0\\0\end{bmatrix}\right)\right\} = \operatorname{Span}\{1\},$$

and

$$\operatorname{Im} T = C_D^{-1}(\mathbb{R}^3) = \mathcal{P}_2(\mathbb{R}).$$

**Exercise 6.6.** Choose the same basis for $\mathcal{P}_2(\mathbb{R})$ and $\mathcal{P}_3(\mathbb{R})$ as above and let $S : \mathcal{P}_2(\mathbb{R}) \to \mathcal{P}_3(\mathbb{R})$ be the linear map defined by letting $S(p)$ be the antiderivative of $p$ with constant term zero. Find the matrix $M_S$ for $S$ and check that $M_T C_B(p) = C_D T(p)$ for all $p \in \mathcal{P}_2(\mathbb{R})$.

**Exercise 6.7.** Suppose $U$, $V$ and $W$ are vector spaces (over the same field) with ordered bases $B$, $D$, and $E$ respectively. Suppose we have linear maps

$$U \xrightarrow{T} V \xrightarrow{S} W.$$

Show that $M_{ST}^{EB} = M_S^{ED} \cdot M_T^{DB}$ (where the product on the right hand side is matrix multiplication). Note that we must use the same basis for the intermediate space $V$ in both matrices.

Note that if we have a linear map $T : V \to V$ from some vector space $V$ to itself, we often use the same bases for $V$ as the domain and as the codomain. But this is not always the case.

**Example 6.8.** Consider the bases $B = \{e_1, e_2\}$ and $D = \{(1,1), (1,-1)\}$ of $\mathbb{R}^2$ and let $I : \mathbb{R}^2 \to \mathbb{R}^2$ be the identity mapping. Then

$$M_I^{DB} = \begin{bmatrix} C_D I(e_1) & C_D I(e_2)\end{bmatrix} = \begin{bmatrix} C_D(e_1) & C_D(e_2)\end{bmatrix}.$$

To finish the calculation, we need to find $C_D(e_1)$ and $C_D(e_2)$. In other words, we need to write $e_1$ and $e_2$ in the basis $D$. We have

$$e_1 = (1,0) = \frac{1}{2}(1,1) + \frac{1}{2}(1,-1), \quad e_2 = \frac{1}{2}(1,1) - \frac{1}{2}(1,-1).$$

Thus

$$M_I^{DB} = \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}.$$

6.2. **Change of bases and similar matrices.** Here we specialize to linear maps $T : V \to V$ from a vector space to itself and assume $V$ is finite-dimensional. What is the relationship between the matrices of $T$ in different bases? If $B$ is an ordered basis of $V$, we'll write $M_T^B$ for $M_T^{BB}$. Recall that

$$M_T^B = C_B T C_B^{-1}, \quad \text{and so} \quad C_B^{-1} M_T^B C_B = T.$$

Now suppose $D$ is another ordered basis of $V$. Then

$$C_B^{-1} M_T^B C_B = T = C_D^{-1} M_T^D C_D,$$

and so

$$M_T^D = C_D C_B^{-1} M_T^B C_B C_D^{-1}.$$

Let $P = C_B C_D^{-1}$. Then $P^{-1} = C_D C_B^{-1}$ and so

$$M_T^D = P^{-1} M_T^B P.$$

Now, if $B = \{v_1, \dots, v_n\}$ and $D = \{w_1, \dots, w_m\}$, then

$$C_D(w_j) = e_j, \quad \text{and so} \quad w_j = C_D^{-1}(e_j).$$

Therefore

$$Pe_j = C_B C_D^{-1}(e_j) = C_B(w_j).$$

Since $Pe_j$ is just the $j$-th column of $P$, we see that

$$\boxed{\text{The } j\text{-th column of } P \text{ gives the coordinates of } w_j \text{ in the basis } B.}$$

Thus, if $P = [P_{ij}]$, then $w_j = \sum_{i=1}^n P_{ij} v_i$. The matrix $P$ is called the *change of basis matrix* from $B$ to $D$.

**Lecture 19: November 25, 2010**

**Example 6.9.** Suppose $V = \mathbb{R}^2$ and $T : \mathbb{R}^2 \to \mathbb{R}^2$ is the linear map given by multiplication by the matrix

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

Let $B = \{e_1, e_2\}$ be the standard basis and $D = \{w_1, w_2\}$ where $w_1 = (1, 1)$ and $w_2 = (1, -1)$. Then

$$M_T^B = \begin{bmatrix} C_B T(e_1) & C_B T(e_2) \end{bmatrix} = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

is the standard matrix of $T$ since $B$ is the standard basis of $\mathbb{R}^2$. However,

$$M_T^D = \begin{bmatrix} C_D\left(T\begin{bmatrix}1\\1\end{bmatrix}\right) & C_D\left(T\begin{bmatrix}1\\-1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} C_D\left(\begin{bmatrix}3\\3\end{bmatrix}\right) & C_D\left(\begin{bmatrix}1\\-1\end{bmatrix}\right) \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix},$$

a diagonal matrix!

Note that $Pe_j = C_B C_D^{-1}(e_j)$, so

$$Pe_1 = C_B C_D^{-1}(e_1) = C_B(w_1) = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad Pe_2 = C_B C_D^{-1}(e_2) = C_B(w_2) = \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

Thus

$$P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix},$$

and

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

Thus

$$P^{-1} M_T^B P = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}\begin{bmatrix} 3 & 1 \\ 3 & -1 \end{bmatrix} = \frac{1}{2}\begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix} = M_T^D.$$

*Remark* 6.10. In this particular case, $w_1$ and $w_2$ are eigenvectors of the matrix $\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$, with eigenvalues 2 and 1, respectively.

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

is diagonal and the columns of $P$ are $w_1$ and $w_2$ (which are $C_B(w_1)$ and $C_B(w_2)$ since $B$ is the standard basis!), and the diagonal entries of $M_T^D$ are the eigenvalues. The whole idea of changing bases is to **simplify** $M_T^B$ as much as possible.

**Exercise 6.11.** Let

$$A = \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}, \quad B = \{e_1, e_2\}, \quad D = \{(3,1), (-2,1)\}, \quad T(v) = Av.$$

Find $M_T^D$ (and show that $A$ is **not** diagonalizable).

**Definition 6.12** (Similar matrices)**.** Two $n \times n$ matrices $X$ and $Y$ are *similar* if there is an invertible matrix $P$ such that $P^{-1}XP = Y$.

**Examples 6.13.**

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ is similar to } \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}.$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ is similar to } \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}, \quad P = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}.$$

$$\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \text{ is similar to } P^{-1}\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}P \text{ and } Q\begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}Q^{-1} \text{ for any } P \text{ (or } Q).$$

*Remark* 6.14. (a) Similarity is an equivalence relation on matrices (**exercise**).

(b) Over an infinite field, a given matrix is **usually** similar to infinitely many matrices. However, there are exceptions. For example, any matrix similar to the identity matrix $I_n$ is of the form $P^{-1}I_nP = P^{-1}P = I_n$. So $I_n$ is only similar to itself. The same is true of $cI_n$ for any scalar $c$.

The next theorem explains why we use the word 'transpose' for the map defined earlier.

**Theorem 6.15.** *Suppose $T : V \to W$ is a linear map between finite-dimensional vector spaces and $T^* : W^* \to V^*$ is the transpose map. If $B$ and $D$ are ordered basis of $V$ and $W$ respectively and $B^*$ and $D^*$ are the corresponding dual bases, then*

$$\left(M_T^{DB}\right)^t = M_{T^*}^{B^*D^*},$$

*where $A^t$ denotes the transpose of the matrix $A$ (as learned in MAT 1341).*

*Proof.* Let

$$B = \{v_1, \ldots, v_n\}, \quad D = \{w_1, \ldots, w_m\}, \quad X = M_T^{DB} = [X_{ij}],$$
$$B^* = \{f_1, \ldots, f_n\}, \quad D^* = \{g_1, \ldots, g_m\}, \quad Y = M_{T^*}^{B^*D^*} = [Y_{ij}].$$

Then, for all $1 \le i \le n$, $1 \le j \le m$, we have

$$(T^*g_j)(v_i) = (g_jT)(v_i) = g_j(Tv_i) = g_j\left(\sum_{k=1}^{m} X_{ki}w_k\right) = X_{ji}.$$

On the other hand,

$$(T^*g_j)(v_i) = \left(\sum_{l=1}^{n} Y_{lj}f_l\right)(v_i) = Y_{ij}.$$

Thus $X_{ji} = Y_{ij}$ for all $1 \le i \le n$ and $1 \le j \le m$. Therefore $X^t = Y$. □

Let's do one final example on matrices associated to linear maps. Suppose one knows the matrix for a linear map in some bases. How do you compute the action of the linear map?

**Example 6.16.** Choose the ordered basis $B = \{1, t, t^2\}$ of $\mathcal{P}_2(\mathbb{R})$ and $D = \{e_{11}, e_{12}, e_{21}.e_{22}\}$ of $M_2(\mathbb{R})$ (here $e_{ij}$ is the matrix with a one in the $(i, j)$-position and zeroes everywhere else). Suppose a linear map $T : \mathcal{P}_2(\mathbb{R}) \to M_2(\mathbb{R})$ has matrix

$$M_T^{BD} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 5 & -1 \end{bmatrix}.$$

What is $T(t^2 - 1)$?

Recall that $M_T^{BD} = C_D T C_B^{-1}$. Thus $T = C_D^{-1} M_T^{BD} C_B$. So

$$\begin{aligned} T(t^2 - 1) &= C_D^{-1} M_T^{BD} C_B(t^2 - 1) \\ &= C_D^{-1} M_T^{BD} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= C_D^{-1} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 5 & -1 \end{bmatrix} \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \\ &= C_D^{-1} \begin{bmatrix} -1 \\ 1 \\ -1 \\ -2 \end{bmatrix} \\ &= -e_{11} + e_{12} - e_{21} - 2e_{22} \\ &= \begin{bmatrix} -1 & 1 \\ -1 & -2 \end{bmatrix} \end{aligned}$$

## 7. Inner product spaces

### 7.1. Definitions.

**Definition 7.1** (Inner product space)**.** A *(real) inner product space* is a vector space over $\mathbb{R}$ together with a map

$$\langle \cdot, \cdot \rangle : V \times V \to \mathbb{R}, \quad (x, y) \to \langle x, y \rangle$$

satisfying

(a) $\langle v, v \rangle \geq 0$ and $\langle v, v \rangle = 0 \iff v = 0$ for all $v \in V$,
(b) $\langle v, w \rangle = \langle w, v \rangle$ for all $v, w \in V$, and
(c) $\langle c_1 v_1 + c_2 v_2, w \rangle = c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle$ for all $v_1, v_2, w \in V$ and $c_1, c_2 \in \mathbb{R}$.

The real number $\langle v, w \rangle$ is called the *inner product of $v$ and $w$*.

*Remark* 7.2. Properties (b) and (c) imply that

$$\langle v, c_1 w_1 + c_2 w_2 \rangle = c_1 \langle v, w_1 \rangle + c_2 \langle v, w_2 \rangle$$

for all $v, w_1, w_2 \in V$ and $c_1, c_2 \in \mathbb{R}$.

**Example 7.3.** On $\mathbb{R}^n$,

$$\begin{aligned} \langle x, y \rangle &:= x \cdot y \qquad \text{(dot product)} \\ &= x^t y \qquad \text{(matrix product)} \end{aligned}$$

is an inner product.

**Example 7.4.** On $C[a, b] := \{f : [a, b] \to \mathbb{R} \mid f \text{ is continuous}\}$,

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

is an inner product. Property (a) is the only difficult property to check. It relies on the fact that if $h$ is a continuous real-valued function on $[a, b]$ and $h(t) \geq 0$ for all $t \in [a, b]$, then

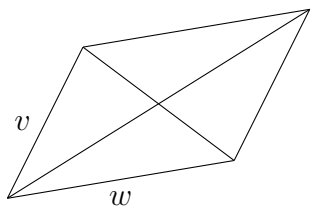
$$\int_a^b h(t)\, dt = 0 \iff h(t) = 0 \ \forall\ t \in [a, b].$$

**Definition 7.5** (Norm)**.** If $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, then the *norm* of $v \in V$ is defined to be

$$\|v\| := \sqrt{\langle v, v \rangle}.$$

**Theorem 7.6.** *If $V$ is an inner product space, then*

*(a)* $\|v\| = 0 \iff v = 0$ *for all* $v \in V$,
*(b)* $\|cv\| = |c|\, \|v\|$ *for all* $c \in \mathbb{R}$ *and* $v \in V$,
*(c)* $\langle v, w \rangle = \frac{1}{4} \left( \|v + w\|^2 - \|v - w\|^2 \right)$ *for all* $v, w \in V$ *(*polarization*), and*
*(d)* $\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2$ *for all* $v, w \in V$ *(parallelogram law).*



*Proof.* The proof of these statements is left as an exercise. *Hint:* Replace the norms by inner products. For the last two parts, expand both sides of each equation. □

**Theorem 7.7** (Cauchy-Schwartz Inequality)**.** *If $(V, \langle \cdot, \cdot \rangle)$ is an inner product space, then*

$$|\langle v, w \rangle| \leq \|v\| \, \|w\| \quad \forall \, v, w, \in V,$$

*and equality holds above if and only if $\{v, w\}$ is linearly dependent.*

*Proof.* Let

$$p(t) = \langle v + tw, v + tw \rangle, \quad t \in \mathbb{R}.$$

Then $p(t) \geq 0$ for all $t \in \mathbb{R}$. Now

$$p(t) = \langle v, v \rangle + t\langle v, w \rangle + t\langle w, v \rangle + t^2 \langle w, w \rangle = \|v\|^2 + 2t\langle v, w \rangle + t^2 \|w\|^2,$$

and so $p$ is a quadratic polynomial. Since $p(t) \geq 0$ for all $t \in \mathbb{R}$, it can have at most one root. Recall that if a polynomial $p(t) = at^2 + bt + c$ has at most one root, then $b^2 - 4ac \leq 0$ (with equality if and only if $p$ has exactly one root). Thus

$$b^2 - 4ac = 4\langle v, w \rangle^2 - 4\|v\|^2 \|w\|^2 \leq 0 \implies |\langle v, w \rangle| \leq \|v\| \|w\|$$

and equality holds it and only if there exists a **unique** $t_0 \in \mathbb{R}$ such that

$$0 = p(t_0) = \langle v + t_0 w, v + t_0 w \rangle,$$

which implies that $v + t_0 w = 0$, and so $\{v, w\}$ is linearly independent. □

## Lecture 20: November 29, 2010

**Announcement:** Some minor adjustments have been made to Assignment 6: point values have been added for each question and a small clarification has been added to Question 4 (the scalars should not be all zero).

**Example 7.8.** In the setting of Example 7.4, we have

$$\left| \int_a^b f(t) g(t) \, dt \right| \leq \sqrt{\int_a^b f(t)^2 \, dt \cdot \int_a^b g(t)^2 \, dt}.$$

This is not an obvious fact at all (try proving it directly using methods from calculus – it's hard)!

**Corollary 7.9.** *If $V$ is an inner product space, then for all $v, w \in V$, we have*

$$\|v + w\| \leq \|v\| + \|w\|.$$

*This is called the* triangle inequality*.*

*Proof.* We have

$$\|v + w\|^2 = \langle v + w, v + w \rangle = \|v\|^2 + 2\langle v, w \rangle + \|w\|^2 \leq \|v\|^2 + 2\|v\| \, \|w\| + \|w\|^2 = (\|v\| + \|w\|)^2.$$

Taking square roots gives the triangle inequality. □

7.2. **Orthogonality.**

**Definition 7.10** (Orthogonal and orthonormal)**.** If $V$ is an inner product space and $v, w \in V$, we say $v$ and $w$ are *orthogonal* if $\langle v, w\rangle = 0$. We say a set $\{v_1, \dots, v_n\}$ of vectors in $V$ is orthogonal if

$$\langle v_i, v_j\rangle = \begin{cases} 0, & i \neq j, \\ \|v_i\|^2 \neq 0, & i = j. \end{cases}$$

In other words, they are nonzero and pairwise orthogonal. We say the set $\{u_1, \dots, u_m\}$ is *orthonormal* if it is orthogonal and $\|u_i\|^2 = 1$ for all $i = 1, \dots, m$ (i.e. each $u_i$ is a *unit vector*).

**Theorem 7.11.** *If $\{v_1, \dots, v_n\}$ is orthogonal, then it is linearly independent.*

*Proof.* Suppose

$$c_1 v_1 + \cdots + c_n v_n = \mathbf{0}$$

for some scalars $c_1, \dots, c_n$. Then for each $j = 1, \dots, n$,

$$0 = \langle \mathbf{0}, v_j\rangle = \left\langle \sum_{i=1}^{n} c_i v_i, v_j \right\rangle = \sum_{i=1}^{n} c_i \langle v_i, v_j\rangle = c_j \|v_j\|^2.$$

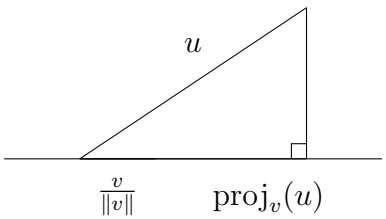
Since $\|v_j\|^2 \neq 0$, this implies $c_j = 0$. Since this is true for all $j = 1, \dots, n$, we see that the set is linearly independent. □

**Definition 7.12** (Projection)**.** Suppose $V$ is an inner product space. For $v \in V$, $v \neq 0$, define the *projection* map $\operatorname{proj}_v : V \to V$ by

$$\operatorname{proj}_v u = \frac{\langle u, v\rangle}{\|v\|^2} v.$$

Note

$$\operatorname{proj}_v u = \left\langle u, \frac{v}{\|v\|} \right\rangle \frac{v}{\|v\|}.$$




**Theorem 7.13.** *If $U$ is a finite-dimensional subspace of an inner product space $V$, then $U$ has an orthogonal basis (hence an orthonormal basis as well).*

The proof of this theorem involves a process called the *Gram-Schmidt algorithm*.

*Proof.* Let $\{v_1, \dots, v_k\}$ be any basis of $U$. We produce an orthogonal basis $\{w_1, \dots, w_k\}$ of $U$ as follows.

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \operatorname{proj}_{w_1} v_2 \\ w_3 &= v_3 - \operatorname{proj}_{w_1} v_3 - \operatorname{proj}_{w_2} v_3 \\ &\vdots \\ w_k &= v_k - \sum_{l=1}^{k-1} \operatorname{proj}_{w_l} v_k. \end{aligned}$$

We claim that $\{w_1, \dots, w_k\}$ is orthogonal. Then it follows from Theorem 7.11 that it is independent, and hence a basis of $U$ (since $U$ has dimension $k$).

We prove the claim by induction. For $1 \leq n \leq k$, let $P(n)$ be the assertion that

(a) $\operatorname{Span}\{w_1, \dots, w_n\} = \operatorname{Span}\{u_1, \dots, u_n\}$, and
(b) $\{w_1, \dots, w_n\}$ is orthogonal.

We know that $P(1)$ is true because $\operatorname{Span}\{w_1\} = \operatorname{Span}\{v_1\}$ (since $w_1 = v_1$) and $\{w_1\} = \{v_1\}$ is orthogonal since $w_1 = v_1 \neq 0$.

Now we show that, for $1 \leq n < k$, $P(n) \implies P(n+1)$. So assume $P(n)$ is true. In other words, $\operatorname{Span}\{w_1, \dots, w_n\} = \operatorname{Span}\{v_1, \dots, v_n\}$, and $\{w_1, \dots, w_n\}$ are orthogonal. Then, for $1 \leq i \leq n$, we have

$$\begin{aligned} \langle w_{n+1}, w_i \rangle &= \langle v_{n+1}, w_i \rangle - \left\langle \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l \rangle}{\|w_l\|^2} w_l, w_i \right\rangle \\ &= \langle v_{n+1}, w_i \rangle - \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l \rangle}{\|w_l\|^2} \langle w_l, w_i \rangle \\ &= \langle v_{n+1}, w_i \rangle - \frac{\langle v_{n+1}, w_i \rangle}{\|w_i\|^2} \langle w_i, w_i \rangle \\ &= 0. \end{aligned}$$

Moreover, if $w_{n+1} = 0$, then

$$v_{n+1} = \sum_{l=1}^{n} \frac{\langle v_{n+1}, w_l \rangle}{\|w_l\|^2} w_l \in \operatorname{Span}\{w_1, \dots, w_n\},$$

and so, by the induction hypothesis $P(n)$(a), $v_{n+1} \in \operatorname{Span}\{v_1, \dots, v_n\}$. But this is impossible since $\{v_1, \dots, v_{n+1}\}$ is linearly independent. Thus $w_{n+1} \neq 0$ and so $\{w_1, \dots, w_{n+1}\}$ is orthogonal. So $P(n+1)$(b) holds.

Now, since

$$w_{n+1} \in \operatorname{Span}\{w_1, \dots, w_n, v_{n+1}\} = \operatorname{Span}\{v_1, \dots, v_{n+1}\},$$

and $P(n)$(a) holds, we know

$$\operatorname{Span}\{w_1, \dots, w_{n+1}\} \subseteq \operatorname{Span}\{v_1, \dots, v_{n+1}\}.$$

Moreover,

$$v_{n+1} = w_{n+1} - \sum_{l=1}^{n} \operatorname{proj}_{w_l} v_k \in \operatorname{Span}\{w_1, \ldots, w_n, w_{n+1}\},$$

so that

$$\operatorname{Span}\{v_1, \ldots, v_{n+1}\} \subseteq \operatorname{Span}\{w_1, \ldots, w_{n+1}\}.$$

Thus $P(n+1)$(a) also holds. This completes the proof by induction.

We can then form an orthonormal basis

$$\left\{ \frac{w_1}{\|w_1\|}, \ldots, \frac{w_k}{\|w_k\|} \right\}.$$

□

**Example 7.14.** Take $V = \mathcal{P}_2(\mathbb{R})$ with the inner product

$$\langle p, q \rangle = \int_{-1}^{1} p(t)q(t)\, dt, \quad p, q \in \mathcal{P}_2(\mathbb{R}).$$

Note that, after restricting the domain of the polynomial functions to $[-1, 1]$, $\mathcal{P}_2(\mathbb{R})$ is a subspace of the inner product space $C[-1, 1]$ (since polynomials are continuous) and the above is the restriction of the inner product on $C[-1, 1]$ to $\mathcal{P}_2(\mathbb{R})$.

Let $\{1, t, t^2\}$ be the standard basis of $\mathcal{P}_2(\mathbb{R})$. We'll find an orthonormal basis using the Gram-Schmidt algorithm. We have

$$w_1 = 1,$$
$$w_2 = t - \operatorname{proj}_1 t = t - \frac{\langle t, 1 \rangle}{\|1\|^2} \cdot 1$$

Now,

$$\langle t, 1 \rangle = \int_{-1}^{1} t \cdot 1\, dt = \frac{1}{2} t^2 \Big|_{-1}^{1} = \frac{1}{2} - \frac{1}{2} = 0.$$

Therefore,

$$w_2 = t - 0 = t$$
$$w_3 = t^2 - \frac{\langle t^2, 1 \rangle}{\|1\|^2} 1 - \frac{\langle t^2, t \rangle}{\|t\|^2} t.$$

Since

$$\langle t^2, 1 \rangle = \int_{-1}^{1} t^2\, dt = \frac{1}{3} t^3 \Big|_{-1}^{1} = \frac{1}{3} - \frac{-1}{3} = \frac{2}{3},$$
$$\|1\|^2 = \langle 1, 1 \rangle = \int_{-1}^{1} 1 \cdot 1\, dt = 2,$$
$$\langle t^2, t \rangle = \int_{-1}^{1} t^3\, dt = \frac{1}{4} t^4 \Big|_{-1}^{1} = 0.$$

Thus,

$$w_3 = t^2 - \frac{2}{3} \cdot \frac{1}{2} \cdot 1 = t^2 - \frac{1}{3}.$$

Hence,

$$\left\{1, t, t^2 - \frac{1}{3}\right\}$$

is an orthogonal basis of $\mathcal{P}_2(\mathbb{R})$.

Now, if we wanted an orthonormal basis of $\mathcal{P}_2(\mathbb{R})$, we compute

$$\|1\| = \sqrt{2}, \quad \|t\|^2 = \int_{-1}^{1} t^2\, dt = \frac{2}{3} \implies \|t\| = \frac{\sqrt{6}}{3},$$

and

$$\|w_3\| = \int_{-1}^{1} \left(t^2 - \frac{1}{3}\right)^2 dt = \int_{-1}^{1} \left(t^4 - \frac{2}{3}t^2 + \frac{1}{9}\right) dt = \frac{1}{5}t^2 - \frac{2}{9}t^3 + \frac{1}{9}t \Big|_{-1}^{1} = \frac{2}{5} - \frac{4}{9} + \frac{2}{9} = \frac{8}{45}.$$

Let

$$u_1 = \frac{w_1}{\|w_1\|} = \frac{\sqrt{2}}{2}, \quad u_2 = \frac{w_2}{\|w_2\|} = \frac{\sqrt{6}}{2}t, \quad u_3 = \frac{w_3}{\|w_3\|} = \frac{3\sqrt{10}}{4}\left(t^2 - \frac{1}{3}\right).$$

Then $\langle u_i, u_j \rangle = \delta_{ij}$.

**Lecture 21: December 2, 2010**

**Final Exam:**

- 8 questions (some have multiple parts).
- Covers material from entire course, but material towards the end of the course is slightly emphasized (since you've already been tested on the earlier material on the midterm).
- Old exams posted on course website. Note that there is no bonus question on this year's final exam (unlike on some old final exams).
- Best way to study is to do recommended exercises. Also, you should go over your old assignments and your midterm and make sure you understand your mistakes.

Orthogonal bases are practical, since if $\{u_1, \dots, u_k\}$ is an orthogonal basis for $U$, and $v \in U$, then

$$v = \frac{\langle v, u_1 \rangle}{\|u_1\|^2} u_1 + \frac{\langle v, u_2 \rangle}{\|u_2\|^2} u_2 + \cdots + \frac{\langle v, u_k \rangle}{\|u_k\|^2} u_k.$$

In other words,

$$v = \sum_{i=1}^{k} c_i u_i, \quad \text{where} \quad c_i = \frac{\langle v, u_i \rangle}{\|u_i\|^2}$$

are the *Fourier coefficients* of $v$ with respect to $u_1, \dots, u_k$. This is convenient, since it saves us time. We don't need to solve a linear system for the coefficients $c_1, \dots, c_n$ as we normally would need to. Computing Fourier coefficients is usually less work (once the orthogonal basis is known).

**Definition 7.15** (Orthogonal matrix)**.** A matrix $A \in M_n(F)$ is called *orthogonal* if $A^t A = I$.

**Lemma 7.16.** *A matrix $A \in M_n(\mathbb{R})$ is orthogonal if and only if its columns form an orthonormal basis of $\mathbb{R}^n$ (with the dot product as the inner product).*

*Proof.* Let $A = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ (i.e. $v_i$ is the $i$-th column of $A$). Then

$$A^t A = \begin{bmatrix} v_1^t \\ \vdots \\ v_n^t \end{bmatrix} \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}.$$

So the $(i,j)$ entry of $A^t A$ is $v_i^t v_j$. Thus

$$A^t A = I_n \iff v_i^t v_j = \delta_{ij} \iff \{v_1, \ldots, v_n\} \text{ is orthonormal.}$$

Since $F^n$ is $n$-dimensional and orthonormal sets are linearly independent, the result follows. □

**Definition 7.17** (Orthogonal complement)**.** Let $U$ be a subspace of an inner product space $V$. Then

$$U^\perp \stackrel{\text{def}}{=} \{v \in V \mid \langle v, u \rangle = 0 \ \forall \ u \in U\}$$

is called the *orthogonal complement* to $U$. It is read "$U$ perp".

**Theorem 7.18.** *Suppose $V$ is a finite-dimensional inner product space and $U$ is a subspace of $V$. Then*

*(a) $U^\perp$ is also a subspace of $U$, and*
*(b) $V = U \oplus U^\perp$.*

The proof will use the important idea of "duality" in inner product spaces.

**Lemma 7.19.** *Suppose $V$ is an inner product space. Define $\varphi : V \to V^*$ by*

$$\varphi(v)(w) = \langle v, w \rangle.$$

*If $V$ is finite-dimensional, then $\varphi$ is an isomorphism.*

*Proof.* Suppose $c_1, c_2 \in \mathbb{R}$ and $v_1, v_2 \in V$. Then for all $w \in V$,

$$\begin{aligned} \varphi(c_1 v_1 + c_2 v_2)(w) &= \langle c_1 v_1 + c_2 v_2, w \rangle \\ &= c_1 \langle v_1, w \rangle + c_2 \langle v_2, w \rangle \\ &= (c_1 \varphi(v_1))(w) + c_2 \varphi(v_2)(w). \end{aligned}$$

Thus

$$\varphi(c_1 v_1 + c_2 v_2) = c_1 \varphi(v_1) + c_2 \varphi(v_2),$$

and so $\varphi$ is linear. Since $\dim V = \dim V^*$, it suffices to show that $\varphi$ is injective. Suppose $\varphi(v) = 0$ for some $v \in V$. Then, in particular,

$$0 = \varphi(v)(v) = \langle v, v \rangle \|v\|^2,$$

and so $v = 0$. Hence $\varphi$ is injective and therefore an isomorphism. □

**Corollary 7.20** (Riesz Representation Theorem)**.** *If $V$ is an inner product space, then for all $f \in V^*$, there exists a unique $v \in V$ such that*

$$f(w) = \langle v, w \rangle, \quad \forall \ w \in V.$$

*In other words, all linear forms are given by taking the inner product with some fixed vector of $V$.*

*Proof.* This follows from the fact that $\varphi$ in Lemma 7.19 is bijective. □

*Proof of Theorem 7.18.* The proof of Part (a) is left as an exercise. We will prove Part (b). So we need to show that $U \cap U^\perp = \{\mathbf{0}\}$ and $U + U^\perp = V$. Suppose $v \in U \cap U^\perp$. Then

$$\langle v, u\rangle = 0 \quad \forall\ u \in U,$$

since $v \in U^\perp$. But since we also have $v \in U$, we have $\langle v, v\rangle = 0$. Thus $v = \mathbf{0}$. So $U \cap U^\perp = \{\mathbf{0}\}$.

Now we show $U + U^\perp = V$. Let $v \in V$ and $f := \varphi(v) \in V^*$ which satisfies

$$f(w) = \langle v, w\rangle \quad \forall\ w \in V.$$

Define $g : U \to \mathbb{R}$ by

$$g(u) = f(u) \quad \forall\ u \in U.$$

Then $g \in U^*$. Now, $U$ is a finite-dimensional inner product space as well (we simply restrict the inner product on $V$ to $U$). Therefore, by the Riesz Representation Theorem for $U$, there exists a $\tilde{u} \in U$ such that

$$g(u) = \langle \tilde{u}, u\rangle \quad \forall\ u \in U.$$

Therefore,

$$g(u) = f(u)\ \forall\ u \in U \implies \langle \tilde{u}, u\rangle = \langle v, u\rangle\ \forall\ u \in U \implies \langle v - \tilde{u}, u\rangle = 0\ \forall\ u \in U \implies v - \tilde{u} \in U^\perp.$$

So

$$v = \tilde{u} + (v - \tilde{u}) \in U + U^\perp.$$

Since $v$ was arbitrary, we have $V = U + U^\perp$. □

**Corollary 7.21.** *If $U$ is a subspace of a finite-dimensional inner product space $V$, then*

$$\dim U^\perp = \dim V - \dim U.$$

**Corollary 7.22.** *Recall that if $U$ is a subspace of an inner product space $V$, then $V = U \oplus U^\perp$. Define*

$$\operatorname{proj}_U : V \to V, \quad \operatorname{proj}_U(u + x) = u, \quad u \in U, \quad x \in U^\perp.$$

*This map has the following properties:*

*(a)* $\operatorname{Im} \operatorname{proj}_U = U$,
*(b)* $\operatorname{Ker} \operatorname{proj}_U = U^\perp$,
*(c)* $\langle v - \operatorname{proj}_U v, u\rangle - 0$ *for all* $vin V$,
*(d)* $\|v - \operatorname{proj}_U v\| \le \|v - u\|$ *for all* $u \in U$, *and equality holds if and only if* $u = \operatorname{proj}_U v$ *(equivalently, if and only if* $v \in U$*).*

*Proof.* Statements (a) and (b) are easy. If $v = u + x = \operatorname{proj}_U v + x$ (for $u \in U$, $x \in U^\perp$), then $v - \operatorname{proj}_U v = x \in U^\perp$, so (c) holds.

It remains to show (d). Let $\tilde{u} = \operatorname{proj}_U v$, so that $v = \tilde{u} + x$, with $x \in U^\perp$. Now let $u$ be **any** vector in $U$. Then

$$\begin{aligned}
\|v - u\|^2 &= \|v - \tilde{u} + \tilde{u} - u\| \\
&= \|x + (\tilde{u} - u)\|^2 \\
&= \|x\|^2 + 2\langle x, \tilde{u} - u\rangle + \|\tilde{u} - u\|^2 \\
&= \|x\|^2 + \|\tilde{u} - u\|^2.
\end{aligned}$$

Thus

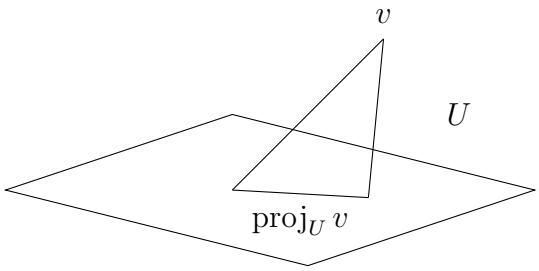
$$\|v - u\|^2 = \|v - \tilde{u}\|^2 + \|\tilde{u} - u\|^2$$

and so

$$\|v - u\|^2 \geq \|v - \tilde{u}\|^2.$$

Moreover, equality holds if and only if $\|\tilde{u} - u\|^2 = 0$, i.e. $u = \tilde{u} = \operatorname{proj}_U v$. □

The map $\operatorname{proj}_U$ is called *orthogonal projection* to $U$. The vector $\operatorname{proj}_U v$ is the "best approximation" to $v$ by a vector in $U$.




*Remark* 7.23. If $\{u_1, \ldots, u_k\}$ is any orthogonal basis for $U$, then an explicit formula for $\operatorname{proj}_U$ is

$$\operatorname{proj}_U v = \frac{\langle v, u_1\rangle}{\|u_1\|^2} u_1 + \frac{\langle v, u_2\rangle}{\|u_2\|^2} u_2 + \cdots + \frac{\langle v, u_k\rangle}{\|u_k\|^2} u_k.$$

Indeed, if $v'$ denotes the formula on the right side, then $v' \in U$, and a short computation shows that $v - v' \in U^\perp$. By uniqueness of the decomposition $v = \tilde{u} + x$, $\tilde{u} \in U$, $x \in U^\perp$, we conclude that $v' = \operatorname{proj}_U v$.

**Corollary 7.24.** *If $U$ is a subspace of a finite-dimensional inner product space $V$, then $\left(U^\perp\right)^\perp = U$.*

*Proof.* If $u \in U$, then $\langle u, v\rangle = 0$ for all $v \in U^\perp$. Thus, by definition, $u \in \left(U^\perp\right)^\perp$. So $U \subseteq \left(U^\perp\right)^\perp$. Moreover,

$$\dim \left(U^\perp\right)^\perp = \dim V - \dim U^\perp = \dim V - (\dim V - \dim U) = \dim U.$$

Thus $U = \left(U^\perp\right)^\perp$. □

**Corollary 7.25.** *The isomorphism $\varphi : U \oplus U^\perp = V \to V^*$ satisfies $\varphi\left(U^\perp\right) = U^0$.*

*Proof.* We have

$$\varphi(v) \in U^0 \iff \varphi(v)(u) = 0 \ \forall\ u \in U \iff \langle v, u\rangle = 0 \ \forall\ u \in U \iff v \in U^\perp.$$

□

7.3. **Duality and adjoints.** Recall that if $V$ is an inner product space, then we have an isomorphism

$$\varphi_V : V \to V^*, \quad \varphi(v)(w) = \langle v, w\rangle, \quad \forall v, w \in V. \tag{7.1}$$

**Definition 7.26** (Adjoint of a linear map)**.** If $T : V \to W$ is a linear map on a finite-dimensional inner product space, the *adjoint* of $T$ is the linear map $T^\star : W \to V$ defined by

$$T^\star = \varphi_V^{-1} \circ T^* \circ \varphi_W,$$

where $T^* : W^* \to V^*$ is the transpose or dual map of $T$.

$$\begin{array}{ccc} V & \xleftarrow{T^\star} & W \\ {\scriptstyle \varphi_V^{-1}}\uparrow & & \downarrow{\scriptstyle \varphi_W} \\ V^* & \xleftarrow[T^*]{} & W^* \end{array}$$

**Proposition 7.27.** *If $T : V \to V$ is a linear map on a finite-dimensional inner product space, then*

$$\langle Tv, w\rangle = \langle v, T^\star w\rangle \quad \forall\ v, w, \in V. \tag{7.2}$$

*Proof.* Let $\varphi = \varphi_V : V \to V^*$ be isomorphism of (7.1). Then $T^\star = \varphi^{-1} \circ T^* \circ \varphi$ and so $\varphi \circ T^\star = T^* \circ \varphi$. Now, for all $v, w \in V$, we have

$$(\varphi \circ T^\star)(w)(v) = (\varphi(T^\star w))(v) = \langle T^\star w, v\rangle = \langle v, T^\star w\rangle,$$

while

$$(T^* \circ \varphi)(w)(v) = T^*(\varphi(w))(v) = \varphi(w)(Tv) = \langle w, Tv\rangle = \langle Tv, w\rangle.$$

Thus

$$\langle Tv, w\rangle = \langle v, T^\star w\rangle \quad \forall\ v, w, \in V.$$

□

**Proposition 7.28.** *Property (7.2) characterizes the adjoint. In others, if $S : V \to V$ satisfies*

$$\langle Tv, w\rangle = \langle v, Sw\rangle \quad \forall\ v, w \in V,$$

*Then $S = T^\star$.*

*Proof.* We have

$$\begin{aligned} \langle v, Sw\rangle = \langle v, T^\star w\rangle\ \forall\ v, w \in V &\iff \langle v, (S - T^\star)w\rangle = 0\ \forall\ v, w \in V \\ &\iff (S - T^*)w = 0\ \forall\ w \in V \iff S = T^\star. \end{aligned}$$

□

*Remark* 7.29. In the text $T'$ denotes the transpose of $T$ and $T^*$ denotes the adjoint of $T$. So make sure you keep this straight when doing the exercises in the text (in particular, $T^*$ denotes the transpose in the class notes but the adjoint in the text).

Often in the literature, the same symbol is used for the transpose and the adjoint. This is because the isomorphism $\varphi : V \to V^*$ is so 'natural' for an inner product space that we 'erase' the it in the equation $\varphi \circ T^\star = T^* \circ \varphi$. Moreover, as we will see next, in the classic example of $\mathbb{R}^n$ with the dot product, the adjoint corresponds to the transpose of a matrix.

**Lecture 22: December 6, 2010**

**Some of the exam questions are recommended exercises.**

**Remember that Wednesday is a Monday schedule: class at 8:30am.**

**Topics for review on Wednesday:**

- dual spaces,
- quotients,

- partitions and equivalence classes,
- isomorphisms, First Isomorphism Theorem.

**Theorem 7.30.** *Fix $A = M_n(\mathbb{R})$ (i.e. $A$ is an $n \times n$ matrix with real entries). Suppose $T : \mathbb{R}^n \to \mathbb{R}^n$ is defined by $T(v) = Av$ for all $v \in \mathbb{R}^n$. Then if $\mathbb{R}^n$ is given the dot product, we have*

$$T^\star(v) = A^t v \quad \forall\ v \in \mathbb{R}^n.$$

*Proof.* Let $v, w \in \mathbb{R}^n$. Then

$$\begin{aligned} \langle Tv, w\rangle &= (Av) \cdot w && \text{(dot product)} \\ &= (Av)^t w && \text{(matrix product)} \\ &= (v^t A^t) w \\ &= v^t A^t w \\ &= v \cdot (A^t w) \\ &= \langle v, A^t w\rangle. \end{aligned}$$

Therefore, by Proposition 7.28, $T^\star(v) = A^t v$ for all $v \in \mathbb{R}^n$. □

## 8. Symmetric matrices and diagonalization

Recall that a matrix $A$ is *symmetric* if $A = A^t$. In the context of $\mathbb{R}^n$ with the dot product, these correspond to linear maps $T : \mathbb{R}^n \to \mathbb{R}^n$ satisfying $T = T^\star$.

**Definition 8.1** (Self-adjoint)**.** Suppose $V$ is an inner product space. A linear map $T : V \to V$ is *self-adjoint* if

$$T = T^\star.$$

**Example 8.2.** Consider $(\mathbb{R}^n, \text{dot product})$. Then $T : \mathbb{R}^n \to \mathbb{R}^n$ defined by $T(v) = Av$, $v \in \mathbb{R}^n$, is self-adjoint if and only if $A = A^t$.

**Lemma 8.3.** *If $B$ is an orthonormal basis of a finite-dimensional inner product space $V$, and $C_B : V \to \mathbb{R}^n$ is the coordinate isomorphism, then*

*(a) $\langle v, w\rangle = C_B(v) \cdot C_B(w)$ (dot product), and*
*(b) if $T : V \to V$ is linear, then $M_{T^\star}^B = \left(M_T^B\right)^t$. So $T$ is self-adjoint if and only if $M_T^B$ is symmetric.*

*Proof.* Let $B = \{w_1, \dots, w_n\}$ and $v, w \in V$. Denote

$$C_B(v) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix}, \qquad C_B(w) = \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix}.$$

Then

$$\begin{aligned} \langle v, w\rangle &= \left\langle \sum_{i=1}^n a_i w_i, \sum_{j=1}^n b_j w_j \right\rangle \\ &= \sum_{i,j=1}^n a_i b_j \langle w_i, w_j\rangle \\ &= \sum_{i=1}^n a_i b_i \\ &= \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} \cdot \begin{bmatrix} b_1 \\ \vdots \\ b_n \end{bmatrix} \\ &= C_B(v) \cdot C_B(w). \end{aligned}$$

It remains to prove the second statement. For all $v, w \in V$, we have

$$\begin{aligned} \langle Tv, w\rangle &= C_B(Tv) \cdot C_B(w) = \left(M_T^B C_B(v)\right) \cdot C_B(w) \\ &= \left(M_T^B C_B(v)\right)^t C_B(w) = C_B(v)^t \left(M_T^B\right)^t C_B(w) \\ &= C_B(v) \cdot \left(\left(M_T^B\right)^t C_B(w)\right), \end{aligned}$$

and

$$\langle v, T^\star w\rangle = C_B(v) \cdot C_B(T^\star w).$$

Since this holds for all $v$, we have

$$x \cdot \left( \left(M_T^B\right)^t C_B(w) \right) = x \cdot C_B(T^\star w) \quad \forall\, x \in \mathbb{R}^n \text{ and } \forall\, w \in V.$$

Therefore

$$\left(M_T^B\right)^t C_B(w) = C_B(T^\star w) \quad \forall\, w \in V,$$

and so

$$\left(M_T^B\right)^t C_B = C_B T^\star,$$

which implies

$$\left(M_T^B\right)^t = M_{T^\star}^B.$$

□

*Remark* 8.4. Sometimes the notation $A^T$ is used for the transpose of the matrix $A$ (instead of $A^t$).

**Corollary 8.5.** *If $D$ is an orthonormal basis of $V$, and $\{y_1, \ldots, y_n\}$ is an orthonormal basis of $\mathbb{R}^n$ (with the dot product), then $B = \{C_D^{-1}(y_1), \ldots, C_D^{-1}(y_n)\}$ is also an orthonormal basis of $V$.*

*Proof.* By Lemma 8.3(a), we have

$$\langle C_D^{-1}(y_i), C_D^{-1}(y_j) \rangle = y_i \cdot y_j = \delta_{ij}.$$

□

**Theorem 8.6.** *If $V$ is a finite-dimensional inner product space and $T : V \to V$ is **self-adjoint**, then there is an orthonormal basis of $V$ consisting of eigenvectors of $T$. That is, there exists an orthonormal basis $B = \{v_1, \ldots, v_n\}$, such that for all $1 \le i \le n$, we have $T(v_i) = \lambda_i v_i$ for some $\lambda_1, \ldots, \lambda_n \in \mathbb{R}$.*

*Remark* 8.7. There is a basis $B$ of eigenvectors of $T$ if and only if $M_T^B$ is a diagonal matrix. (You saw this in MAT 1341 in the case $V = \mathbb{R}^n$ and $T$ is multiplication by a matrix.) Indeed if $B = \{v_1, \ldots, v_n\}$, then

$$\begin{aligned}
Tv_i = \lambda_i v_i \ \forall\ i = 1, \ldots, n &\iff C_B(Tv_i) = C_B(\lambda_i v_i) \ \forall\ i = 1, \ldots, n \\
&\iff M_T^B C_B(v_i) = \lambda_i e_i \ \forall\ i = 1, \ldots, n \\
&\iff M_T^B e_i = \lambda_i e_i \ \forall\ i = 1, \ldots, n \\
&\iff M_T^B = \begin{bmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \ldots & 0 & \lambda_n \end{bmatrix}.
\end{aligned}$$

**Corollary 8.8.** *If $A \in M_n(\mathbb{R})$ is symmetric, then there exists an orthogonal matrix $C$ and a diagonal matrix $D$ such that $C^T A C = D$.*

**Example 8.9.** Consider the matrix

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

Find a diagonal matrix $D$ and an orthogonal matrix $C$ such that $C^T A C = D$.

First we find the eigenvalues of $A$. We set

$$0 = \det(A - \lambda I) = \begin{vmatrix} 6-\lambda & 4 \\ 4 & -1-\lambda \end{vmatrix} = (5-\lambda)(-1-\lambda) - 16 = \lambda^2 - 4\lambda - 21 = (\lambda - 7)(\lambda + 3).$$

Thus the eigenvalues are 7 and $-3$. For the eigenvalue $\lambda = -3$, we find the eigenspace $E_{-1}$ by solving

$$(A - \lambda I)x = \mathbf{0} \implies (A + 3I)x = \mathbf{0}.$$

We do this by row reducing

$$\left[\begin{array}{cc|c} 8 & 4 & 0 \\ 4 & 2 & 0 \end{array}\right] \rightsquigarrow \left[\begin{array}{cc|c} 2 & 1 & 0 \\ 0 & 0 & 0 \end{array}\right],$$

and so

$$E_{-3} = \operatorname{Span}\left\{\begin{bmatrix} 1 \\ -2 \end{bmatrix}\right\}.$$

Since we want $C$ to be an orthogonal matrix, its columns must form an orthonormal basis (of eigenvectors). So we need a unit vector in $E_{-1}$. Take

$$v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ -2 \end{bmatrix}.$$

Now, we could repeat this process with the eigenvalue 7. However, we can save ourselves some effort by using Theorem 8.6 to conclude that $A$ (which is symmetric) must have an orthonormal basis. So we take

$$v_2 = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

(we're using here the fact that $(-b, a)$ is orthogonal to $(a, b)$). Then $\{v_1, v_2\}$ is an orthonormal basis of eigenvectors,

$$C = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix}$$

is an orthogonal matrix, and

$$C^T A C = \frac{1}{5}\begin{bmatrix} 1 & -2 \\ 2 & 1 \end{bmatrix}\begin{bmatrix} 5 & 4 \\ 4 & -1 \end{bmatrix}\begin{bmatrix} 1 & 2 \\ -2 & 1 \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 0 & 7 \end{bmatrix}.$$

**Example 8.10.** Let $T : \mathbb{R}^3 \to \mathbb{R}^3$ be defined by $Tv = Av$, $v \in \mathbb{R}^3$, where

$$A = \begin{bmatrix} 3 & 2 & 4 \\ 2 & 0 & 2 \\ 4 & 2 & 3 \end{bmatrix}.$$

Let's find an orthogonal matrix $C$ such that $C^T A C = D$ is diagonal.

First we find the eigenvalues of $A$.

$$\det(A - \lambda I) = \begin{vmatrix} 3-\lambda & 2 & 4 \\ 2 & -\lambda & 2 \\ 4 & 2 & 3-\lambda \end{vmatrix} \overset{\text{exercise}}{=} -(\lambda + 1)^2(\lambda - 8).$$

So the eigenvalues are $-1$ (with multiplicity 2) and 8 (with multiplicity 1). Now,

$$E_{-1} = \mathrm{Ker}(A+I) = \mathrm{Ker}\begin{bmatrix} 4 & 2 & 4 \\ 2 & 1 & 2 \\ 4 & 2 & 4 \end{bmatrix} = \mathrm{Ker}\begin{bmatrix} 2 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{Span}\left\{ \begin{bmatrix} -1/2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\}.$$

We convert the basis of the eigenspace into an orthogonal basis using the Gram-Schmidt algorithm:

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \right\} \rightsquigarrow \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} - \frac{1}{5}\begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix} \right\} = \left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} -4/5 \\ -2/5 \\ 1 \end{bmatrix} \right\}$$

or (since we can scale each vector by anything we like)

$$\left\{ \begin{bmatrix} -1 \\ 2 \\ 0 \end{bmatrix}, \begin{bmatrix} 4 \\ 2 \\ -5 \end{bmatrix} \right\}$$

Similarly,

$$E_8 = \mathrm{Ker}(A-8I) = \mathrm{Ker}\begin{bmatrix} -5 & 2 & 4 \\ 2 & -8 & 2 \\ 4 & 2 & -5 \end{bmatrix} = \mathrm{Ker}\begin{bmatrix} 1 & 0 & -1 \\ 0 & 2 & -1 \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{Span}\left\{ \begin{bmatrix} 1 \\ 1/2 \\ 1 \end{bmatrix} \right\}.$$

Thus

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

is a basis for $E_8$.

We now form the matrix $C$ we're after by using the basis vectors we've found (after dividing by their norm in order to make them unit vectors) as the columns. Thus

$$C = \begin{bmatrix} -\sqrt{5}/5 & 4\sqrt{5}/15 & 2/3 \\ 2\sqrt{5}/5 & 2\sqrt{5}/15 & 1/3 \\ 0 & -\sqrt{5}/3 & 2/3, \end{bmatrix}$$

and

$$C^T AC = D = \begin{bmatrix} -1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 8 \end{bmatrix}.$$

*Remark* 8.11. The only thing new here (over what you did in MAT 1341) is that we're dealing with symmetric matrices and we're not looking for just **any** basis of eigenvectors, but rather an **orthonormal** basis of eigenvectors. So in eigenspaces of dimension greater than 1, we must use the Gram-Schmidt algorithm to get an orthonormal basis of the eigenspace. Eigenvectors corresponding to different eigenvalues are **automatically** orthogonal (when the matrix is symmetric).

**Lecture 23: December 8, 2010**

## 9. Review

We do not have time to review the entire course. Instead we'll review the topics suggested by students. You should **not** assume that topics not covered in this review will not be on the exam. The exam covers the **entire** course, whether or not it is mentioned in this review.

9.1. **Dual spaces.** If $V$ is a vector space over the field $F$, the *dual space* of $V$ is

$$V^* = \mathcal{L}(V, F).$$

So elements of the dual space are linear maps from $V$ to $F$.

If $\{v_1, \dots, v_n\}$ is a basis of $V$, then $\{f_1, \dots, f_n\}$ where $f_i$, $1 \leq i \leq n$, is defined by

$$f_i(v_j) = \delta_{ij}, \quad 1 \leq i, j \leq n,$$

is the *dual basis* of $V^*$.

If $T : V \to W$ is a linear map, then its *transpose* is the linear map $T^* : W^* \to V^*$ defined by

$$T(f) = f \circ T, \quad f \in W^*.$$

We picture this as

$$\begin{array}{ccc} V & \xrightarrow{T} & W \\ & \searrow^{f \circ T} & \downarrow f \\ & & F \end{array}.$$

9.2. **Equivalence relations and partitions.** If $X$ is any set and $\sim$ is a relation on $X$ (i.e. it is a property where, for any $x, y \in A$, the statement $x \sim y$ is either true or false), then $\sim$ is an *equivalence relation* if

- $x \sim x$ for all $x \in X$,
- for all $x, y \in X$, $x \sim y \implies y \sim x$,
- for all $x, y, z \in X$, ($x \sim y$ and $y \sim z$) $\implies y \sim z$.

**Example 9.1.** If $X = \mathbb{Z}$, then

$$x \sim y \iff x - y \in 2\mathbb{Z}, \quad x, y \in X,$$

is an equivalence relation on $X$.

A *partition* of a set $X$ is a collection $\mathcal{A}$ of subsets of $X$ such that

- every $A \in \mathcal{A}$ is a nonempty subset of $X$,
- if $A, B \in \mathcal{A}$ and $A \neq B$, then $A \cap B = \varnothing$,
- every element of $X$ belongs to one of the subsets in $\mathcal{A}$.

**Example 9.2.** If $X = \mathbb{Z}$, then $\mathcal{A} = \{2\mathbb{Z}, 2\mathbb{Z} + 1\}$ is a partition of $X$.

If $\sim$ is an equivalence relation on $X$ and $x \in X$, then

$$[x] = \{y \in X \mid y \sim x\}$$

is the *equivalence class* of $x$. The set of equivalence classes of any equivalence relation on $X$ is a partition of $X$.

**Example 9.3.** With the equivalence relation of Example 9.1, the set of equivalence classes is the partition of Example 9.2.

9.3. **Quotients.** If $M$ is a subspace of a vector space $V$, then we can define an equivalence relation on $V$ by

$$u \sim v \iff u - v \in M, \quad u, v \in V.$$

The equivalence class of $u \in V$ is

$$[u] = u + M = \{u + x \mid x \in M\}.$$

We define

$$V/M = \{u + M \mid u \in V\}$$

to be the set of equivalence classes and call it the *quotient* of $V$ by $M$. By the above, $V/M$ is also a partition of $V$.

We have *quotient map*

$$Q : V \to V/M, \quad Q(v) = v + M.$$

Some important properties of the quotient map are:

- $Q$ is linear,
- $Q$ is surjective,
- $\operatorname{Ker} Q = M$.

9.4. **Isomorphisms and the First Isomorphism Theorem.** An *isomorphism* is a bijective linear map. The *First Isomorphism Theorem* says that if $T : V \to W$ is a linear map, then

$$V/\operatorname{Ker} T \cong T(V) = \operatorname{Im} T.$$

**Example 9.4.** Let

$$V = \{p \in \mathcal{P}(\mathbb{R}) \mid p(0) = 0\}$$

Consider the map

$$T : \mathcal{P}(\mathbb{R}) \to \mathbb{R}, \quad T(p) = p(0).$$

Then $\operatorname{Ker} T = V$ and so $V$ is a subspace of $\mathcal{P}(\mathbb{R})$. Also, $T$ is surjective since for any $x \in \mathbb{R}$, if $p$ is the constant polynomial given by $p(t) = x$, then $T(p) = x$. Thus, by the First Isomorphism Theorem,

$$\mathcal{P}(\mathbb{R})/V \cong \mathbb{R}.$$

Remember that if $T$ is a linear map, you can show it is injective by showing that $\operatorname{Ker} T = \{0\}$.

**Example 9.5.** Let

$$S = \{(x_1, x_2, x_3, \dots) \mid x_i \in \mathbb{R} \ \forall\ i = 1, 2, 3, \dots\}$$

be the vector space of all infinite sequences of real numbers. Show that the map

$$T : \mathcal{P}(\mathbb{R}) \to S, \quad T(p) = (p(1), p(2), p(3), \dots)$$

is injective.

Since $T$ is linear (exercise), we can show it is injective by showing that $\operatorname{Ker} T = \{0\}$. Suppose $p \in \operatorname{Ker} T$. Then

$$T(p) = 0 \implies (p(1), p(2), p(3), \dots) = (0, 0, 0, \dots) \implies p(i) = 0 \ \forall\ i = 1, 2, 3, \dots.$$

This means that $p$ has infinitely many roots. But any nonzero polynomial has a finite number of roots (less than or equal to the degree of the polynomial). So $p = 0$. Thus $\operatorname{Ker} T = \{0\}$.

9.5. **Some recommended exercises from the text.**

**Section 5.3, Exercise 1.** Let $T : \mathbb{R}^3 \to \mathbb{R}^2$ be the linear mapping defined by

$$T(a, b, c) = (2a - c, 3b + 4c).$$

(i) Viewing $\mathbb{R}^3$ and $\mathbb{R}^2$ as inner product spaces (with the dot product), find the matrix of $T^\star$ relative to the canonical orthonormal bases, then deduce a formula for $T^\star$.

First note that $T$ is given by multiplication by the matrix

$$\begin{bmatrix} Te_1 & Te_2 & Te_3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & -1 \\ 0 & 3 & 4 \end{bmatrix}$$

We know (Theorem 7.30) that $T^\star$ is given by multiplication by the matrix

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

Therefore

$$T^\star(a, b) = \begin{bmatrix} 2 & 0 \\ 0 & 3 \\ -1 & 4 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} = \begin{bmatrix} 2a \\ 3b \\ -a + 4b \end{bmatrix} = (2a, 3b, -a + 4b).$$

(ii) Find the matrix of $T^\star$ relative to the basis $B = \{\frac{1}{2}(1, \sqrt{3}), \frac{1}{2}(-\sqrt{3}, 1)\}$ of $\mathbb{R}^2$ and the canonical basis $D = \{e_1, e_2, e_3\}$ of $\mathbb{R}^3$.

We have

$$\begin{aligned} M_{T^\star}^{DB} &= \begin{bmatrix} C_D\left(T\left(\frac{1}{2}(1, \sqrt{3})\right)\right) & C_D\left(T\left(\frac{1}{2}(-\sqrt{3}, 1)\right)\right) \end{bmatrix} \\ &= \begin{bmatrix} C_D\left(1, \frac{3\sqrt{3}}{2}, -\frac{1}{2} + 2\sqrt{3}\right) & C_D\left(-\sqrt{3}, \frac{3}{2}, \frac{\sqrt{3}}{2} + 2\right) \end{bmatrix} \\ &= \begin{bmatrix} 1 & -\sqrt{3} \\ 3\sqrt{3}2 & 3/2 \\ -1/2 + 2\sqrt{3} & \frac{\sqrt{3}}{2} + 2 \end{bmatrix}. \end{aligned}$$

**Section 5.2, Exercise 6.** If $M$ and $N$ are linear subspaces of a Euclidean space (finite dimensional inner product space), then $(M + N)^\perp = M^\perp \cap N^\perp$ and $(M \cap N)^\perp = M^\perp + N^\perp$.

Suppose $x \in (M + N)^\perp$. Since $M, N \subset M + N$, we have $x \in M^\perp \cap N^\perp$. So $(M + N)^\perp \subseteq M^\perp \cap N^\perp$. Now suppose $x \in M^\perp \cap N^\perp$. Then, for any $y \in M$, $z \in N$, we have

$$\langle x, y + z \rangle = \langle x, y \rangle + \langle x, z \rangle = 0 + 0 = 0.$$

So $x \in (M + N)^\perp$. Thus $M^\perp + N^\perp \subseteq (M + N)^\perp$. Hence $M^\perp + N^\perp = (M + N)^\perp$.

Now, $M^\perp$ and $N^\perp$ are also subspaces of the same finite-dimensional inner product space. So we can replace $M$ by $M^\perp$ and $N$ by $N^\perp$ in the equation we just proved. This gives us

$$(M^\perp)^\perp + (N^\perp)^\perp = (M^\perp + N^\perp)^\perp \implies M + N = (M^\perp + N^\perp)^\perp.$$

(Here we use the fact that for a **finite-dimensional** inner product space, we have $(U^\perp)^\perp = U$ for any subspace $U$.) Taking the orthogonal complement of both sides then gives

$$(M + N)^\perp = \left((M^\perp + N^\perp)^\perp\right)^\perp = M^\perp + N^\perp,$$

as desired.

# Index