MTH 309

Supplemental Lecture Notes

Based on Robert Messer, Linear Algebra Gateway to Mathematics

Ulrich Meierfrankenfeld

Department of Mathematics

Michigan State University

East Lansing MI 48824

meier@math.msu.edu

Contents

Chapter N1

Vector Spaces

N1.1 Logic and Sets

In this section we will provide an informal discussion of logic and sets. We start with a little bit of logic:

A statement is a sentence which is either true or false, for example

1. $1 + 1 = 2$
2. $\sqrt{2}$ is a rational number.
3. π is a real number.
4. Exactly 1323 bald eagles were born in 2000 BC,

all are statements. Statement (1) and (3) are true. Statement (2) is false. Statement (4) is probably false, but verification might be impossible. It nevertheless is a statement.

Let P and Q be statements.

"P and Q" is the statement that P is true and Q is true.

"P or Q" is the statement that at least one of P and Q is true.

So "P or Q" is false if both P and Q are false.

"$\neg P$' (pronounced 'not P' or 'negation of P') is the statement that P is false. So $\neg P$ is true if P is false. And $\neq P$ is false if P is true.

"P $\Longrightarrow$ Q" (pronounced "P implies Q") is the statement "$\neq P$ or Q". Note that "P$\Longrightarrow$ Q" is true if P is false. But if P is true, then "P$\Longrightarrow$ Q" is true if and only if Q is true. So one often uses the phrase "If P is true, then Q is true" or "if P, then Q" in place of "P$\Longrightarrow$ Q"

"P$\Longleftrightarrow$ Q" (pronounced "P is equivalent to Q") is the statement "(P and Q) or (not-P and not-Q)". So "P$\Longleftrightarrow$ Q" is true if either both P and Q are true or both P and Q are false. So one often uses the phrase "P holds if and only if Q holds", or "P if and only if Q" in place of "P$\Longleftrightarrow$ Q"

One can summarize the above statements in the following truth table:

P	Q	$\neg P$	$\neg Q$	P and Q	P or Q	$P \Longrightarrow Q$	$P \Longleftrightarrow Q$
T	T	F	F	T	T	T	T
T	F	F	T	F	T	F	F
F	T	T	F	F	T	T	F
F	F	T	T	F	F	T	T

In the following we collect a few statements which are always true.

Lemma N1.1.1. *Let P, Q and R be statements, let T be true statement and F a false statement. Then each of the following statements holds.*

LR 1 $F \Longrightarrow P$.

LR 2 $P \Longrightarrow T$.

LR 3 $\neg(\neg P) \Longleftrightarrow P$.

LR 4 $(\neg P \Longrightarrow F) \Longrightarrow P$.

LR 5 P or T.

LR 6 $\neg(P \text{ and } F)$.

LR 7 $(P \text{ and } T) \Longleftrightarrow P$.

LR 8 $(P \text{ or } F) \Longleftrightarrow P$.

LR 9 $(P \text{ and } P) \Longleftrightarrow P$.

LR 10 $(P \text{ or } P) \Longleftrightarrow P$.

LR 11 P or $\neg P$.

LR 12 $\neg(P \text{ and } \neg P)$.

LR 13 $(P \text{ and } Q) \Longleftrightarrow (Q \text{ and } P)$.

LR 14 $(P \text{ or } Q) \Longleftrightarrow (Q \text{ or } P)$.

LR 15 $(P \Longleftrightarrow Q) \Longleftrightarrow \Big((P \text{ and } Q) \text{ or } (\neg P \text{ and } \neg Q)\Big)$

LR 16 $(P \Longrightarrow Q) \Longleftrightarrow (\neg P \text{ or } Q)$.

LR 17 $\neg(P \Longrightarrow Q) \Longleftrightarrow (P \text{ and } \neg Q)$.

LR 18 $\Big(P \text{ and} (P \implies Q)\Big) \implies Q.$

LR 19 $\Big((P \implies Q) \text{ and} (Q \implies P)\Big) \iff (P \iff Q).$

LR 20 $(P \implies Q) \iff (\neg Q \implies \neg P)$

LR 21 $(P \iff Q) \iff (\neg P \iff \neg Q).$

LR 22 $\neg(P \text{ and } Q) \iff (\neg P \text{ or } \neg Q)$

LR 23 $\neg(P \text{ or } Q) \iff (\neg P \text{ and } \neg Q)$

LR 24 $\Big((P \text{ and } Q) \text{ and } R\Big) \iff \Big(P \text{ and} (Q \text{ and } R)\Big).$

LR 25 $\Big((P \text{ or } Q) \text{ or } R\Big) \iff \Big(P \text{ or} (Q \text{ or } R)\Big).$

LR 26 $\Big((P \text{ and } Q) \text{ or } R\Big) \iff \Big((P \text{ or } R) \text{ and} (Q \text{ or } R)\Big).$

LR 27 $\Big(P \text{ or } Q) \text{ and } R\Big) \iff \Big((P \text{ and } R) \text{ or} (Q \text{ and } R)\Big).$

LR 28 $\Big((P \implies Q) \text{ and} (Q \implies R)\Big) \implies (P \implies R)$

LR 29 $\Big((P \iff Q) \text{ and} (Q \iff R)\Big) \implies (P \iff R)$

Proof. If any of these statements are not evident to you, you should use a truth table to verify it. □

The contrapositive of the statement $P \implies Q$ is the statements $\neg Q \implies \neg P$. (LR 20) says the contrapositive $\neg Q \implies \neg P$ is equivalent to $P \implies Q$. Indeed, both are equivalent to $P \text{ and} (\neg Q)$.

The contrapositive of the statement $P \iff Q$ is the statements $\neg P \iff \neg Q$. (LR 21) says the contrapositive $\neg P \iff \neg Q$ is equivalent to $P \iff Q$.

The converse of the implication $P \implies Q$ is the statement $Q \implies P$. The converse of an implication is not equivalent to the original implication. For example the statement if $x = 0$ then x is an even integer is true. But the converse (if x is an even integer, then $x = 0$) is not true.

Theorem N1.1.2 (Principal of Substitution)**.** *Let $\Phi(x)$ be a formula involving a variable x. If a and b are objects with $a = b$, then $\Phi(a) = \Phi(b)$.*

Proof. This should be self evident. For an actual proof and the definition of a formula consult your favorite logic book. □

We now will have a short look at sets.

First of all any *set* is a collection of objects.

For example

$$\mathbb{Z} := \{\ldots, -4, -3, -2, -1, -0, 1, 2, 3, 4, \ldots\}$$

is the set of integers. If S is a set and x an object we write $x \in S$ if x is a member of S and $x \notin S$ if x is not a member of S. In particular,

$$(*) \qquad \text{For all } x \text{ exactly one of} \quad x \in S \quad \text{and} \quad x \notin S \quad \text{holds.}$$

Not all collections of objects are sets. Suppose for example that the collection $\mathcal{B}$ of all sets is a set. Then $\mathcal{B} \in \mathcal{B}$. This is rather strange, but by itself not a contradiction. So lets make this example a little bit more complicated. We call a set S is nice, if $S \notin S$. Let $\mathcal{D}$ be the collection of all nice sets and suppose $\mathcal{D}$ is a set.

Is $\mathcal{D}$ a nice?

Suppose that $\mathcal{D}$ is a nice. Since $\mathcal{D}$ is the collection of all nice sets, $\mathcal{D}$ is a member of $\mathcal{D}$. Thus $\mathcal{D} \in \mathcal{D}$, but then by the definition of nice, $\mathcal{D}$ is not nice.

Suppose that $\mathcal{D}$ is not nice. Then by definition of nice, $\mathcal{D} \in \mathcal{D}$. Since $\mathcal{D}$ is the collection of nice sets, this means that $\mathcal{D}$ is nice.

We proved that $\mathcal{D}$ is nice if and only if $\mathcal{D}$ is not nice. This of course is absurd. So $\mathcal{D}$ cannot be a set.

Theorem N1.1.3. *Let A and B be sets. Then*

$$(A = B) \iff \Big(\text{for all } x : (x \in A) \iff (x \in B)\Big)$$

Proof. Naively this just says that two sets are equal if and only if they have the same members. In actuality this turns out to be one of the axioms of set theory. □

Definition N1.1.4. *Let A and B be sets. We say that A is* subset *of B and write $A \subseteq B$ if*

$$\text{for all } x : (x \in A) \Longrightarrow (x \in B)$$

In other words, A is a subset of B if all the members of A are also members of B.

Lemma N1.1.5. *Let A and B sets. Then $A = B$ if and only if $A \subseteq B$ and $B \subseteq A$.*

Proof.

$$\begin{array}{rcl} & A = B & \\ \iff & x \in A \iff x \in B & -\text{ N1.1.3} \\ \iff & (x \in A \Longrightarrow x \in B) \text{ and } (x \in B \Longrightarrow x \in A) & -(\text{LR } 19) \\ \iff & A \subseteq B \text{ and } B \subseteq A & -\text{definition of subset} \end{array}$$

□

Theorem N1.1.6. *Let S be a set and let $P(x)$ be a statement involving the variable x. Then there exists a set, denoted by $\{s \in S \mid P(s)\}$ such that*

$$\Big(t \in \{s \in S \mid P(s)\}\Big) \iff \Big(t \in S \text{ and } P(t)\Big)$$

Proof. This follows from the so called replacement axiom in set theory. □

Note that an object t is a member of $\{s \in S \mid P(s)\}$ if and only if t is a member of S and the statement $P(t)$ is true For example

$$\{x \in \mathbb{Z} \mid x^2 = 1\} = \{1, -1\}.$$

Theorem N1.1.7. *Let S be a set and let $\Phi(x)$ be a formula involving the variable x such that $\Phi(s)$ is defined for all s in S. Then there exists a set, denoted by $\{\Phi(s) \mid s \in S\}$ such that*

$$\Big(t \in \{\Phi(s) \mid s \in S\}\Big) \iff \Big(\text{ There exists } s \in S \text{ with } t = \Phi(s)\Big)$$

Proof. This also follows from the replacement axiom in set theory. □

Note that the members of $\{\Phi(s) \mid s \in S\}$ are all the objects of the form $\Phi(s)$, where s is a member of S.

For example $\{2x \mid x \in \mathbb{Z}\}$ is the set of even integers.

We can combined the two previous theorems into one:

Theorem N1.1.8. *Let S be a set, let $P(x)$ be a statement involving the variable x and $\Phi(x)$ a formula such that $\Phi(s)$ is defined for all s in S for which $P(s)$ is true. Then there exists a set, denoted by $\Big\{\Phi(s) \mid s \in S \text{ and } P(s)\Big\}$ such that*

$$\left(t \in \Big\{\Phi(s) \mid s \in S \text{ and } P(s)\Big\}\right) \iff \left(\text{There exists } s \in S \text{ with } \Big(P(s) \text{ and } t = \Phi(s)\Big)\right)$$

Proof. Just define

$$\Big\{\Phi(s) \mid s \in S \text{ and } P(s)\Big\} = \Big\{\Phi(t) \mid t \in \{s \in S \mid \Phi(s)\}\Big\}$$

□

Note that the members of $\{\Phi(s) \mid s \in S \text{ and } P(s)\}$ are all the objects of the form $\Phi(s)$, where s is a member of S for which $P(s)$ is true.

For example

$$\{2n \mid n \in \mathbb{Z} \text{ and } n^2 = 1\} = \{2, -2\}$$

Theorem N1.1.9. *Let A and B be sets.*

(a) There exists a set, denoted by $A \cup B$ and called A union B, such that

$$(x \in A \cup B) \Longleftrightarrow (x \in A \text{ or } x \in B)$$

(b) There exists a set, denoted by $A \cap B$ and called A intersect B, such that

$$(x \in A \cap B) \Longleftrightarrow (x \in A \text{ and } x \in B)$$

(c) There exists a set, denoted by $A \setminus B$ and called A removed B, such that

$$(x \in A \setminus B) \Longleftrightarrow (x \in A \text{ and } x \notin B)$$

(d) There exists a set, denoted by $\emptyset$ and called empty set, such that

$$\textit{For all } x: \qquad x \notin \emptyset$$

Proof. (a) This is another axiom of set theory.

(b) Applying N1.1.6 with $P(x)$ being the statement "$x \in B$" we can define

$$A \cup B = \{x \in A \mid x \in B\}$$

(c) Applying N1.1.6 with $P(x)$ being the statement "$x \notin B$" we can define

$$A \setminus B = \{x \in A \mid x \notin B\}$$

(d) One of the axioms of set theory implies the existence of a set A. Then we can define

$$\emptyset = A \setminus A$$

□

Let A be a set. Since the empty set has no members, all of its members are in A. So

Lemma N1.1.10. *Let A be a set. Then $\emptyset \subseteq A$.*

Proof. Here is a slightly more formal proof: Let x be an object. By definition of the emptyset, $x \notin \emptyset$. Thus the statement $x \in \emptyset$ is false and so by (LR 1) the implication $(x \in \emptyset) \Longrightarrow (x \in A)$ is true. So $\emptyset \subseteq A$ holds by the definition of a subset. □

Lemma N1.1.11. *Let A and B be sets. Then $A \cap B = B \cap A$.*

Proof. Let x be an object. Then

$$\begin{aligned} & x \in A \cap B \\ \Longleftrightarrow\ & x \in A \text{ and } x \in B && - \text{Definition of } A \cap B \\ \Longleftrightarrow\ & x \in B \text{ and } x \in A && -(\text{LR } 14) \\ \Longleftrightarrow\ & x \in B \cap A && - \text{Definition of } A \cap B \end{aligned}$$

So $A \cap B = B \cap A$. □

N1.2 Basic Definition

Definition 1.1. *A* vector space $\mathbf{V}$ *is a triple* $(V, \oplus, \odot)$ *such that*

(i) V *is a set (whose elements are called* vectors*)*

(ii) $\oplus$ *is a function (called* vector addition*) ,* $V \times V$ *is a subset of the domain of* $\oplus$ *and*

$$v \oplus w \in V \qquad \textit{(}\text{Closure of addition}\textit{)}$$

for all $v, w \in V$*, where* $v \oplus w$ *denotes the image of* (v, w) *under* $\oplus$*;*

(iii) $\odot$ *is a function (*called scalar *multiplication),* $\mathbb{R} \times V$ *is a subset of the domain of* $\odot$ *and*

$$r \odot v \in V \qquad \textit{(}\text{Closure of multiplication}\textit{)}$$

for all $r \in \mathbb{R}$ *and* $v \in V$*, where* $r \odot v$ *denotes the image of* (r, v) *under* $\odot$*;*

and such that the following eight statements hold:

(Ax 1) $v \oplus w = w \oplus v$ *for all* $v, w \in V$*;* *(*Commutativity of Addition*)*

(Ax 2) $v \oplus (w \oplus x) = (v \oplus w) \oplus x$ *for all* $v, w, x \in V$*;* *(*Associativity of Addition*)*

(Ax 3) *There exists an element in* V*, denoted by* $\mathbf{0_V}$ *(and called an* additive identity*), such that* $v \oplus 0_{\mathbf{V}} = v$ *for all* $v \in V$*;* (Existence of Additive Identity)

(Ax 4) *For each* $v \in V$ *there exists an element in* V*, denoted by* $-v$ *(and called an additive inverse of* v*), such that* $v \oplus (-v) = \mathbf{0_V}$*;* (Existence of Additive Inverse)

(Ax 5) $a \odot (v \oplus w) = (a \odot v) \oplus (a \odot w)$ *for all* $a \in \mathbb{R}$ *and* $v, w \in V$*;* (Right Distributivity)

(Ax 6) $(a + b) \odot v = (a \odot v) \oplus (b \odot v)$ *for all* $a, b \in \mathbb{R}$*,* $v \in V$*;* (Left Distributivity)

(Ax 7) $(ab) \odot v = a \odot (b \odot v)$ *for all* $a, b \in \mathbb{R}$, $v \in V$; (Associativity of Multiplication)

(Ax 8) $1 \odot v = v$ *for all* $v \in V$. (Multiplicative Identity)

Remark N1.2.2. *Often slightly different version of conditions (ii) and (iii) are used in the definition of a vector space:*

(ii') $\oplus$ *is a function (called vector addition) from* $V \times V$ *to* V *and* $v \oplus w$ *denotes the image of* (v, w) *under* $\odot$;

(iii') $\odot$ *is a function (called scalar multiplication) from* $\mathbb{R} \times V$ *to* V *and* $r \odot v$ *denotes the image of* (r, v) *under* $\odot$;

Note that Conditions (ii') and (iii') imply (ii) and (iii) Conversely, if $\oplus$ *and* $\odot$ *fulfill (ii) and (iii) and one replaces* $\oplus$ *by its restriction to* $V \times V$ *and* V, *and* $\odot$ *by its restriction to* $\mathbb{R} \times V$ *and* V, *then (ii') and (iii') hold. So there is no essential difference between these two definitions of a vector space.*

Notation N1.2.3. *Given a vector space* $\mathbf{V} = (V, \oplus, \odot)$. *We will often use the following simplified notation, where* $r \in \mathbb{R}$ *and* $v, w \in V$:

$$v + w \text{ denotes } v \oplus w$$

$$rv \text{ denotes } r \odot v, \text{ and}$$

$$\mathbf{0} \text{ denotes } \mathbf{0}_{\mathbf{V}}.$$

Observe that we now use the same symbols for the addition and multiplication in $\mathbf{V}$ *as in* $\mathbb{R}$. *So we will use this notation only in situations where it should not lead to confusion.*

N1.3 Properties of Vector Spaces

Lemma N1.3.1. *Let* $\mathbf{V}$ *be vector space and* $v, w \in V$. *Then* $(v + w) + (-w) = v$.

Proof.

$$\begin{aligned} (v+w)+(-w) &= v + (w + (-w)) && -(\text{Ax } 2) \\ &= v + \mathbf{0} && -(\text{Ax } 4) \\ &= v && -(\text{Ax } 3) \end{aligned}$$

□

Theorem 1.8 (Cancellation Law)**.** *Let* V *be vector space and* $v, w, x \in V$. *Then the following three statements are equivalent (that is if one of the three statements holds, all three hold):*

(a) $v = w$.

(b) $x + v = x + w$.

(c) $v + x = w + x$.

Proof. It suffices to show that (a) implies (b), that (b) implies (c) and (c) implies (a). Indeed suppose we proved these three implications. If (a) holds, then since (a) implies (b), also (b) holds; and since (b) implies (c) also (c) holds. Similarly, if (b) holds, then since (b) implies (c), also (c) holds; and since (c) implies (a) also (a) holds. And if (c) holds, then since (c) implies (a), also (a) holds; and since (a) implies (b) also (b) holds. So any of the statements implies the other two.

(a) $\implies$ (b): Suppose that $v = w$. Then $x+v = x+w$ by the Principal of Substitution N1.1.2 and so (a) implies (b) holds.

(b) $\implies$ (c): Suppose that $x + v = x + w$. Then (Ax 1) applied to each side of the equation gives $v + x = w + x$. So (b) implies (c).

(c) $\implies$ (a): Suppose $v + x = x + w$. Adding $-x$ to both sides of the equation gives $(v+x)+(-x) = (w+x)+(-x)$. Applying N1.3.1 to both sides gives $v = w$. So (c) implies (a) and the Cancellation Law holds. □

Theorem 1.2. *Let* $\mathbf{V}$ *be a vector space and* $v, w \in V$. *Then*

(a) $\mathbf{0} + v = v$.

(b) If $v + w = v$, *then* $w = \mathbf{0}$.

(c) If $w + v = v$, *then* $w = \mathbf{0}$.

(d) $\mathbf{0}$ *is the only additive identity in* V.

Proof. (a) By (Ax 1), $\mathbf{0} + v = v + \mathbf{0}$ and by (Ax 3), $v + \mathbf{0} = v$. Thus $\mathbf{0} + v = v$.

(b) Suppose that $v + w = v$. By (Ax 3), $v = v + \mathbf{0}$ and so

$$v + w = v + \mathbf{0}$$

Thus by the Cancellation Law 1.8, $w = \mathbf{0}$.

(b) Suppose that $w + v = v$. Then by (Ax 1), $v + w = v$ and so by (b), $w = \mathbf{0}$.

(d) Let u be an additive identity in V. Then by definition of an additive identity, $\mathbf{0} + u = \mathbf{0}$. By (a) $\mathbf{0} + u = u$ and so $u = z$. □

Theorem 1.3. *Let* $\mathbf{V}$ *be a vector space and* $v, w \in V$. *Then* $v + w = \mathbf{0}$ *if and only if* $w = -v$. *In particular,* $-v$ *is the only additive inverse of* v.

Proof. By (Ax 4), $v + (-v) = \mathbf{0}$. Now suppose that $v + w = \mathbf{0}$. Then $v + w = v + (-v)$ and so by the Cancellation Law 1.8, $w = -v$. So $-v$ is the only additive inverse of v. □

Theorem 1.4. *Let* $\mathbf{V}$ *be a vector space and* $v \in V$. *Then*

$$\mathbf{0}_{\mathbb{R}} v = \mathbf{0}_{\mathbf{V}}$$

Proof. Since $0+0=0$ in $\mathbb{R}$ we have $0v=(0+0)v$. Hence by (Ax 6), $0v=0v+0v$ and so by 1.2(b) (applied to $0v$ and $0v$ in place of v and w), $0v=\mathbf{0}$. $\square$

Theorem 1.5. *Let* $\mathbf{V}$ *be a vector space,* $v\in V$ *and* $r,s\in\mathbb{R}$. *Then*

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

(b) $r\mathbf{0}=\mathbf{0}$.

(c) If $rv=\mathbf{0}$, *then* $r=0$ *or* $v=\mathbf{0}$.

(d) $(-1)v=-v$.

(e) $-v=\mathbf{0}$ *if and only if* $v=\mathbf{0}$.

(f) $-(-v)=v$.

(g) $(-r)v=-(rv)=r(-v)$.

(i) If $v\neq\mathbf{0}$ *and* $rv=sv$, *then* $r=s$.

Proof. For (a), (b), (d) and (i) see Homework 3.

(c) Suppose that $rv=\mathbf{0}$. We need to show that $r=0$ or $v=\mathbf{0}$. If $r=0$, this holds. So we may assume that $r\neq 0$. Then by properties of the real numbers, r has an multiplicative inverse $\frac{1}{r}$. So

$$(*)\qquad \frac{1}{r}\in\mathbb{R}\text{ and }\frac{1}{r}r=1$$

We have

$$\begin{array}{rrcll} & rv & = & \mathbf{0} & -\text{by assumption}\\ \Longrightarrow & \frac{1}{r}(rv) & = & \frac{1}{r}\mathbf{0} & -\text{Principal of Substitution}\\ \Longrightarrow & (\frac{1}{r}r)v & = & \mathbf{0} & -\text{(Ax 7) and Part (b) of the current theorem}\\ \Longrightarrow & 1v & = & \mathbf{0} & -(*)\\ \Longrightarrow & v & = & \mathbf{0} & -\text{(Ax 1)} \end{array}$$

So $v=\mathbf{0}$ and (c) is proved.

(e) We have

$$\begin{array}{rrcll} & \mathbf{0} & = & -v & \\ \Longleftrightarrow & v+\mathbf{0} & = & \mathbf{0} & -\text{Theorem 1.3 applied with } w=\mathbf{0}\\ \Longleftrightarrow & v & = & \mathbf{0} & -\text{(Ax 3)} \end{array}$$

(f) By (Ax 4), $v + (-v) = \mathbf{0}$ and so by (Ax 1), $(-v) + v = \mathbf{0}$. Thus v is an additive inverse of $-v$ and so by 1.3, $v = -(-v)$.

(g) We would like to show that $(-r)v = -(rv)$, that is we would like to show that $(-r)v$ is the additive inverse of rv. We compute

$$\begin{aligned} rv + (-r)v &= (r + (-r))v && - \text{(Ax 6)} \\ &= 0v && - \text{Property of real numbers} \\ &= \mathbf{0} && - \text{Theorem 1.4} \end{aligned}$$

Thus $(-r)v$ is an additive inverse of rv and so by 1.3, $(-r)v = -(rv)$. Hence the first equality in (g) holds. To prove the second we need to verify that also $r(-v)$ is an additive inverse of rv:

$$\begin{aligned} rv + r(-v) &= r(v + (-v)) && -\text{(Ax 5)} \\ &= r\mathbf{0} && -\text{(Ax 4)} \\ &= \mathbf{0} && - \text{Part (b) of the current theorem} \end{aligned}$$

Thus $r(-v)$ is an additive inverse of rv and so by 1.3, $r(-v) = -(rv)$. Hence also the second equality in (g) holds.

□

N1.4 Subtraction

Definition 1.6. *Let* $\mathbf{V}$ *be a vector space and* $v, w \in V$*. Then the vector* $v \ominus w$ *in* V *is defined by*

$$v \ominus w = v \oplus (-w)$$

As long as no confusion should arise, we will just write $v - w$ *for* $v \ominus w$*.*

Theorem 1.7. *Let* $\mathbf{V}$ *be a vector space and* $v, w, x \in V$*. Then each of the following statements holds*

(d) $(-v) - w = (-v) + (-w) = (-w) + (-v) = (-w) - v = -(v + w)$

(g) $(v + w) - w = v$.

(m) $r(v - w) = rv - rw$

(n) $(r - s)v = rv - sv$.

(o) $(v + w) - x = v + (w - x)$.

(p) $v - w = \mathbf{0}$ *if and only if* $v = w$.

(q) $(v-w)+w=v$.

Proof. For the proof of (m) and (n) see Homework 3.

(o)

$$\begin{aligned} (v+w)-x &= (v+w)+(-x) && \text{–Definition of '-', see 1.6} \\ &= v+(w+(-x)) && \text{–(Ax 2)} \\ &= v+(w-x) && \text{–Definition of '-'} \end{aligned}$$

(q)

$$\begin{aligned} (v-w)+w &= (v+(-w))+(-(-w)) && \text{– Definition of '-' and 1.5(f)} \\ &= v && \text{– N1.3.1} \end{aligned}$$

(p)

$$\begin{aligned} && v-w &= \mathbf{0} \\ \iff && (v-w)+w &= \mathbf{0}+w && \text{–Cancellation Law 1.8} \\ \iff && v &= w && \text{–(q) and 1.2} \end{aligned}$$

(g)

$$\begin{aligned} (v+w)-w &= (v+w)+(-w) && \text{–Definition of '-'} \\ &= v && \text{–Lemma N1.3.1} \end{aligned}$$

(d)

$$\begin{aligned} (-v)-w &= (-v)+(-w) && \text{–Definition of '-'} \\ &= (-w)+(-v) && \text{–(Ax 1)} \\ &= (-w)-v && \text{–Definition of '-'} \end{aligned}$$

So the first three equalities in (g) hold. To prove the last, we will show that $-(v+w) = (-w)+(-v)$. For this we need to show that $(-w)+(-v)$ is an additive inverse of $v+w$. We compute

$$\begin{aligned} (v+w)+\Big((-w)+(-v)\Big) &= \Big((v+w)+(-w)\Big)+(-v) && \text{–(Ax 2)} \\ &= v+(-v) && \text{–Lemma N1.3.1} \\ &= \mathbf{0} && \text{–(Ax 4)} \end{aligned}$$

Hence $(-w)+(-v)$ is an additive inverse of $v+w$ and so by 1.3, $(-w)+(-v) = -(v+w)$. So all the elements listed in (d) are equal. $\square$

N1.7 Function Spaces

Definition N1.7.1. *Let I be the set. Then* $\mathrm{F}(I)$ *denotes the set of all functions from I to* $\mathbb{R}$. *For $f, g \in \mathrm{F}(I)$ we define $f+g \in \mathrm{F}(I)$ by*

$$(f+g)(i) = f(i) + g(i)$$

for all $i \in I$. For $r \in \mathbb{R}$ and $f \in \mathrm{F}(i)$ we define $rf \in \mathrm{F}(I)$ by

$$(rf)(i) = r\big(f(i)\big)$$

for all $i \in I$.

$\mathbf{F}(I)$ *is the triple consisting of* $\mathrm{F}(I)$*, the above addition and the above multiplication.*

Theorem N1.7.2. *Let I be a set.*

(a) $\mathbf{F}(I)$ *is a vector space.*

(b) The additive identity *in* $\mathrm{F}(I)$ *is the function* $0^* \in \mathrm{F}(I)$ *defined by* $0^*(i) = 0$ *for all* $i \in I$.

(c) The additive inverse *of $f \in \mathrm{F}(I)$ is the function $-f$ defined by $(-f)(i) = -\big(f(i)\big)$ for all $i \in I$.*

Proof. Properties 1.1(i), (ii) and (iii) hold by definition of $\mathbf{F}(I)$. We will now verify the first four axioms of a vector space one by one. For the remaining four, see Homework 4. From Lemma A.2.2 we have

(*) Let $f, g \in \mathrm{F}(I)$. Then $f = g$ if and only if $f(i) = g(i)$ for all $i \in I$.

Let $f, g, h \in \mathrm{F}(I)$ and $i \in I$.

(Ax 1): We have

$$\begin{aligned}(f+g)(i) &= f(i) + g(i) && \text{—Definition of '+' for functions}\\ &= g(i) + f(i) && \text{—Property of } \mathbb{R}\\ &= (g+f)(i) && \text{—Definition of '+' for functions}\end{aligned}$$

So $f + g = g + f$ by (*) and (Ax 1) is proved.

(Ax 2): We have

$$\begin{aligned}
\Big((f+g)+h\Big)(i) &= (f+g)(i)+h(i) && \text{–Definition of '+' for functions}\\
&= \Big(f(i)+g(i)\Big)+h(i) && \text{–Definition of '+' for functions}\\
&= f(i)+\Big(g(i)+h(i)\Big) && \text{–Property of } \mathbb{R}\\
&= f(i)+\Big(g+h\Big)(i) && \text{–Definition of '+' for functions}\\
&= \Big(f+(g+h)\Big)(i) && \text{–Definition of '+' for functions}
\end{aligned}$$

So $(f+g)+h = f+(g+h)$ by (*) and (Ax 2) is proved.

(Ax 3) Define a function, denoted by 0^*, in $\mathrm{F}(I)$ by $0^*(i) = 0$ for all i. We will show that 0^* is an additive identity:

$$\begin{aligned}
(f+0^*)(i) &= f(i)+0^*(i) && \text{–Definition of '+' for functions}\\
&= f(i)+0 && \text{–Definition of } 0^*\\
&= f(i) && \text{–Property of } \mathbb{R}
\end{aligned}$$

So $f+0^* = f$ by (*) and (Ax 3) is proved.

(Ax 3) Define a function, denoted by $-f$, in $\mathrm{F}(I)$ by $(-f)(i) = -f(i)$ for all i. We will show that $-f$ is an additive inverse of f.

$$\begin{aligned}
\Big(f+(-f)\Big)(i) &= f(i)+(-f)(i) && \text{–Definition of '+' for functions}\\
&= f(i)+(-f(i)) && \text{–Definition of } -f\\
&= 0 && \text{–Property of } \mathbb{R}\\
&= 0^*(i) && \text{–Definition of } 0^*
\end{aligned}$$

So $f+(-f) = 0^*$ by (*) and (Ax 4) is proved. □

N1.5 Euclidean space

Let n be a positive integer, $a_1, \ldots, a_n$ real numbers and $(a_1, a_2, \ldots, a_n)$ the corresponding list of length n. Note that we can view such an list as the function from $\{1, 2, 3, \ldots, n\}$ to $\mathbb{R}$ which maps 1 to a_1, 2 to a_2,..., and n to a_n. In fact the will use this observation to give a precise definition of what we mean with list.

Definition N1.5.1. *Let S be a set and n and m non-negative integers.*

(a) A list of length n in S is a function $f : \{1, \dots, n\} \to S$.

(b) The list of length 0 in S is denoted by $()$ and is called the empty list.

(c) Let $s_1, s_2, \dots s_n \in S$, then $(s_1, s_2, \dots, s_n)$ denotes the unique list f with $f(i) = s_i$ for all $1 \leq i \leq n$.

(d) The list $(t_1, t_2, t_3, \dots, t_m)$ in S is called a sublist of $(s_1, s_2 \dots, s_n)$ if there exist integers $1 \leq i_1 < i_2 < \dots < i_m \leq n$ with $t_j = s_{i_j}$ for all $1 \leq j \leq m$. (In terms of functions: A list g of length m is a sublist of the list f of length n if there exists a strictly increasing function $h : \{1 \dots, m\} \to \{1, \dots, n\}$ with $g = f \circ h$.)

For example $B = (2, 4, 7, 9, 11)$ is a list of length 5 in the integers and $(4, 9, 11)$ is a sublist of B of length 3.

Definition N1.5.2. *Let n be non-negative integer.*

(a) $\mathbb{R}^n$ is the set of all lists of length n in $\mathbb{R}$. So $\mathbb{R}^n = \mathrm{F}(\{1, 2, 3, \dots, n\})$.

(b) $\mathbf{R}^n$ denotes the vector space $\mathbf{F}(\{1, 2, 3, \dots, n\})$.

Lemma N1.5.3. *The* vector addition *and* scalar multiplication *in $\mathbf{R}^n$ can be described as follows: Let $(a_1, a_2, \dots, a_n)$ and $(b_1, b_2, \dots, b_n)$ be elements of $\mathbb{R}^n$ and $r \in \mathbb{R}$. Then*

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

and

$$r(a_1, a_2 \dots, a_n) = (ra_1, ra_2, \dots, ra_n).$$

Proof. Let $f = (a_1, a_2, \dots, a_n)$, $g = (b_1, b_2, \dots b_n)$ and $1 \leq i \leq n$. Then by Definition N1.5.1(c), $f(i) = a_i$ and $g(i) = b_i$. So by the definition of addition on $\mathbf{R}^n$,

$$(f + g)(i) = f(i) + g(i) = a_i + b_i.$$

Hence by Definition N1.5.1(c)

$$f + g = (a_1 + b_1, a_2 + b_2, \dots, a_n + b_n).$$

This gives the first statement.
Also by the definition of multiplication in $\mathbf{R}^n$,

$$(rf)(i) = r\big(f(i)\big) = ra_i,$$

and so Definition N1.5.1(c)

$$rf = (ra_1, ra_2, \dots, ra_n).$$

Thus also the second statement holds. $\square$

N1.6 Matrices

Let n and m be positive integers, let $a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots, a_{2n}, \ldots, a_{m1}, a_{m2} \ldots, a_{mn}$ be real numbers and

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

the corresponding $m \times n$-matrix. Note that we can view such matrix as the function from $\{1, 2, 3, \ldots, m\} \times \{1, 2, \ldots .n\}$ to $\mathbb{R}$ which maps $(1, 1)$ to a_{11}, $(1, 2)$ to a_{12}, …,$(1, n)$ to a_{1n}, $(2, 1)$ to a_{21}, $(2, 2)$ to a_{22}, …, $(2, n)$ to a_{2n},…, $(m, 1)$ to a_{m1}, $(m, 2)$ to a_{m2}, …, and (m, n) to a_{mn}. In fact the will use this observation to give a precise definition of what we mean with an $m \times n$-matrix.

Definition N1.6.1. *Let n, m be positive integers.*

*(a) Let I and J be sets. An $I \times J$-*matrix *is a function from $I \times J$ to $\mathbb{R}$.*

(b) An $m \times n$-matrix is $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$-matrix, that is a function from $\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\}$ to $\mathbb{R}$.

(c) Given real numbers $a_{11}, a_{12}, \ldots, a_{1n}, a_{21}, a_{22}, \ldots a_{2n}, \ldots, a_{m1}, a_{m2} \ldots, a_{mn}$. Then the unique $m \times n$-matrix A with $A(i, j) = a_{ij}$ for all $1 \leq i \leq m$ and $1 \leq j \leq n$ is denoted by

$$\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1n} \\ a_{21} & a_{22} & \ldots & a_{2n} \\ \vdots & \vdots & \ldots & \vdots \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{bmatrix}$$

or by

$$[a_{ij}]_{i=1,j=1}^{m,n}$$

(d) $\mathbb{M}(m, n)$ is the set of all $m \times n$ matrices. So $\mathbb{M}(m, n) = \mathrm{F}(\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\})$

(e) $\mathbf{M}(m, n)$ denotes the vector space $\mathbf{F}(\{1, 2, \ldots, m\} \times \{1, 2, \ldots, n\})$

Lemma N1.6.2. *The vector* addition *and scalar* multiplication *in $\mathbf{M}(m, n)$ can be described as follows: Let $[a_{ij}]_{i=1,j=1}^{m,n}$ and $[b_{ij}]_{i=1,j=1}^{m,n}$ be $m \times n$ matrices and $r \in \mathbb{R}$. Then*

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} + \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & a_{12}+b_{12} & \dots & a_{1n}+b_{1n} \\ a_{21}+b_{21} & a_{22}+b_{22} & \dots & a_{2n}+b_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1}+b_{m1} & a_{m2}+b_{m2} & \dots & a_{mn}+b_{mn} \end{bmatrix}$$

and

$$r\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & \dots & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} ra_{11} & ra_{12} & \dots & ra_{1n} \\ ra_{21} & ra_{22} & \dots & ra_{2n} \\ \vdots & \vdots & \dots & \vdots \\ ra_{m1} & ra_{m2} & \dots & ra_{mn} \end{bmatrix}$$

Proof. Let $A = [a_{ij}]_{i=1,j=1}^{m,n}$ and $B = [b_{ij}]_{i=1,j=1}^{m,n}$. Also let $1 \leq i \leq m$ and $1 \leq j \leq n$. Then by Definition N1.6.1(c), $A(i,j) = a_{ij}$ and $B(i,j) = b_{ij}$. So by the definition of addition in $\mathbf{M}(m,n)$,

$$(A+B)(i,j) = A(i,j) + B(i,j) = a_{ij} + b_{ij}.$$

Hence by Definition N1.6.1(c)

$$A + B = [a_{ij} + b_{ij}]_{i=1,j=1}^{m,n}.$$

This gives the first statement.

Also by the definition of multiplication in $\mathbf{M}(m,n)$,

$$(rA)(i,j) = r\big(A(i,j)\big) = ra_{ij},$$

and so Definition N1.6.1(c)

$$rA = [ra_{ij}]_{i=1,j=1}^{m,n}$$

Thus also the second statement holds. □

N1.8 Subspaces

Definition 1.10. *Let $\mathbf{V} = (V, \oplus, \odot)$ be a vector space and W a subset of V. Put $\mathbf{W} = (W, \oplus, \odot)$. Then $\mathbf{W}$ is called a subspace of $\mathbf{V}$ provided that $\mathbf{W}$ is a vector space.*

Theorem 1.11 (Subspace Theorem)**.** *Let V be a vector space and W a subset of V. Then $\mathbf{W}$ is a subspace of $\mathbf{V}$ if and only if each of the following three statements holds:*

(1) $\mathbf{0_V} \in W$.

(2) $u + w \in W$ *for all* $u, w \in W$.

(3) $rw \in W$ *for all* $r \in \mathbb{R}$ *and* $w \in W$.

Proof. Suppose first that $\mathbf{W}$ is a subspace of $\mathbf{V}$. Condition 1.1(ii) (Closure of addition) for the vector space W shows that $u + w \in W$ for all $u, w \in W$. Similarly, Condition 1.1(iii) (Closure of multiplication) for the vector space W shows that $rw \in W$ for all $r \in \mathbb{R}$, $w \in W$. So (2) and (3) hold. By 1.1(3) W has an additive identity $\mathbf{0_W} \in W$. In particular, $\mathbf{0_W} + \mathbf{0_W} = \mathbf{0_W}$ and so by 1.2

$$(*) \qquad \mathbf{0_W} = \mathbf{0_V}$$

Since $\mathbf{0_W} \in W$, this gives (1).

Suppose next that (1), (2) and (3) hold. We need to show that $\mathbf{W}$ is a vector space.

By assumption W is a subset of V and so Condition 1.1(ii) holds for W.

By 1.1(ii) for V, $\oplus$ is a function and $V \times V$ is contained in the domain of $\oplus$. Since W is subset of V, $W \times W$ is a subset of $V \times V$, and so is contained in the domain of $\oplus$. By (2), $u + w \in W$ for all $u, w \in W$ and so Condition 1.1(ii) holds for W.

By 1.1(iii) for V, $\odot$ is a function and $\mathbb{R} \times V$ is contained in the domain of $\odot$. Since W is subset of V, $W \times W$ is a subset of $V \times V$, and so is contained in the domain of $\oplus$. By (3), $rv \in W$ for all $r \in \mathbb{R}$ and $w \in W$ and so Condition 1.1(ii) holds for W.

Since Axioms 1,2,5,6,7,8 holds for all suitable elements of $\mathbb{R}$ and V, and since $W \subseteq V$, they clearly also hold for all suitable elements in $\mathbb{R}$ and W.

By (1), $\mathbf{0_V} \in W$. Since $\mathbf{0_V}$ is an additive identity for V, we conclude that $\mathbf{0_V}$ is also an additive identity for W in W. So Axioms 3 holds for W with $\mathbf{0_W} = \mathbf{0_V}$.

Let $w \in W$. By 1.5, $-w = (-1)w$. By (3), $(-1)w \in W$ and so

$$(**) \qquad -w \in W$$

Since $w + (-w) = \mathbf{0_V} = \mathbf{0_W}$ we conclude that

$$-w \text{ is the additive inverse of } w \text{ in } \mathbf{W}$$

Hence Axiom 4 holds for W. □

Corollary N1.8.3. *Let* $\mathbf{W}$ *be a subspace of the vector space* $\mathbf{V}$ *and* $w \in W$. *Then* $\mathbf{0_W} = \mathbf{0_V}$, $-w \in W$ *and the additive inverse of* w *in* $\mathbf{W}$ *is the same as the additive inverse of* w *in* $\mathbf{V}$.

Proof. See (*), (**) and (***) in the proof of 1.11 □

Chapter N2

Systems of Linear Equations

N2.1 Notation and Terminology

Definition N2.1.1. *Let $A = (a_{ij})_{i\in I}^{j\in J}$ be an $I \times J$ matrix.*

(a) Let $K \subseteq I$ and $L \subseteq J$. Then $A_L^K = K$ is the restriction of A to $K \times L$, so A_L^K is the $K \times J$- matrix $(a_{ij})_{i\in K}^{j\in L}$. A_L^K is called the $K \times L$ submatrix of A.

(b) Let $i \in I$. Then $\mathbf{a}^i = A_J^{\{i\}}$. $\mathbf{A}^i$ is called Row i of A. a^i is the J-list $(a_{ij})_{j\in J}$. a^i is called row i of A.

(c) Let $j \in J$. Then $\mathbf{a}_j = A_{\{j\}}^I$. $\mathbf{a}_j$ is called Column j of A. a_j is the I-list $(a_{ij})_{i\in I}$. a_j is called column i of A.

(d) Let $i \in I$ and $j \in I$. Then $\mathbf{a}_{ij} = A_{\{j\}}^{\{i\}}$ is called the ij-Entry of A, while a_{ij} is called the ij-entry of A.

Note that a^i and $\mathbf{a}^i$, viewed as functions, have different domains: the domain of $\mathbf{a}_i$ is $\{i\} \times J$, while the domain of a^i is J. On the other hand, the ij-entry of $\mathbf{a}^i$ is the same as the j-entry of $a^i i$ (both are equal to a_{ij}). Informally, Row i knows its position in the matrix, but row i does not.

For example consider $A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \end{bmatrix}$. Then row 1 of A is equal to row 2 of A. But Row 1 of A is not equal to Row 2 of A, since they have different domains. Similarly, the 12-entry of A is equal to the 11-entry of A, but the 12-Entry of A is not equal to the 11-Entry.

Notation N2.1.2. *(a) Let I be a set, n a non-negative integer and for $1 \leq j \leq n$ let a_j be I-list in $\mathbb{R}$. Then*

$$[a_1, a_2, \dots, a_n]$$

denotes the unique $I \times \{1, \dots, n\}$ matrix with column j equal to a_j for all $1 \leq j \leq n$.

(b) Let J be a set, m a non-negative integer and for $1 \leq i \leq m$ let a^i be J-list in $\mathbb{R}$. Then

$$\begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^m \end{bmatrix}$$

denotes the unique $\{1, \ldots, m\} \times J$ matrix with row i equal to a^i for $1 \leq i \leq m$.

Remark: Let K be a set and x a K-list in $\mathbb{R}$. According to (a) $[x]$ denotes the $K \times \{1\}$-matrix whose only column is x, while according to (b), $[x]$ denotes the $\{1\} \times K$ matrix whose only row is a. So the notation $[x]$ is ambiguous and should only be used if it is clear from the context which have the two possible matrices is meant.

Notation N2.1.3. *Let I and J be sets, $i, k \in I$, $A \in M(I, J)$ and $x, y \in F(I)$.*

(a) $R_i x A$ denotes the $I \times J$ matrix B with $b^i = x$ and $b^l = a^l$ for all $l \in I$ with $l \neq i$. (So $R_i x A$ is the matrix obtained from A by replacing Row i be x.

(b) $R_{ik} x y A = R_i x (R_k y A)$. So $R_i x A$ is the matrix obtained from A by replacing Row k be y and then replacing Row i by x.)

Definition 2.1. *Let I and J be sets. An elementary row operation on $M(I, J)$ is one of functions '$R_i \leftrightarrow R_j$', '$cR_i + R_k \to R_k$' and (for $c \neq 0$), '$cR_i \to R_i$' from $M(I, J)$ to $M(I, J)$ defined as below, where $i, k \in I$, $c \in \mathbb{R}$ and A is an $I \times J$ matrix.*

1. *$(R_i \leftrightarrow R_j)A = R_{ik} a^k a^i A$. So $R_i \to R_j$ interchangings row i and row k of A*

2. *$(cR_i + R_k \to R_k)(A) = R_k(ca^i + a^k)A$. So $(cR_i + R_k \to R_k)$ adds c times row i to row k of A.*

3. *Suppose $c \neq 0$. Then $(cR_i \to R_i)A = R_i(ca^i)A$. So $cR_i \to R_i$ multiplies row A of A by c.*

Remark: Column replacements and elementary column operations are defined similarly using the symbol C in place of R.

N2.2 Gaussian Elimination

Definition 2.2. *Let A be an $m \times n$ matrix, $1 \leq i \leq m$ and $1 \leq j \leq n$.*

(a) $\mathbf{a}_{ij}$ is called a leading Entry of A provided that $a_{ij} \neq 0$ and $a_{kj} = 0$ for all $1 \leq k < i$. So a_{ij} is the first non-zero entry in row i of A.

(b) $\mathbf{a}_{ij}$ is called a leading 1 of A if $\mathbf{a}_{ij}$ is a leading Entry and $a_{ij} = 1$.

Algorithm N2.2.2 (Gauss-Jordan)**.** *Let A be an $\{l, l+1, \ldots, m\} \times \{1, \ldots, n\}$-matrix.*

Step 1 *If A is a zero matrix, the algorithm stops. If A is not the zero matrix, let j be minimal such that a_j is not a zero vector. Then let i be minimal such that $a_{ij} \neq 0$. So a_{ij} is the first non-zero entry in the first non-zero column of A.*

Step 2 *Interchange row l and row i of A.*

Step 3 *Multiply row l of A by $\frac{1}{a_{lj}}$ of A.*

Step 4 *For $l < k < m$, add $-a_{kj}$ times row l of A to row k of A.*

Step 5 *Apply the Gauss-Jordan algorithm to the $\{l+1, \ldots, m\} \times \{1, \ldots, n\}$-submatrix of A.*

Definition 2.3. *Let A be an $n \times m$-matrix. Then A is in row-echelon form if*

(i) All leading Entries are leading 1's.

(ii) If $\mathbf{a}_{ij}$ and $\mathbf{a}_{kl}$ are leading 1's with $i < k$, then $j < l$.

(iii) If A_i is a non-zero row and A_j is a zero row, then $i < j$.

Observe that the Gauss-Jordan algorithm produces a matrix in row-echelon form.

Algorithm N2.2.4 (Reduced Gauss-Jordan)**.** *Let A be an $m \times m$ matrix in echelon form.*

Step 6 If A is a zero matrix, the algorithm stops. If A is not the zero matrix, let $\mathbf{a}_{ij}$ be a leading 1 with i maximal. (So $\mathbf{a}_{ij}$ is the leading 1 in the last non-zero row of A.)

Step 7 For $1 \leq k < i$, add $-a_{kj}$ times row i of A to row k of A.

Step 8 Apply the Reduced Gauss-Jordan algorithm to the $\{1, \ldots, i-1\} \times \{1, \ldots, n\}$-submatrix of A.

Definition 2.4. *Let A be an $n \times m$-matrix. Then A is in reduced row-echelon form if*

(i) All leading Entries are leading 1's.

(ii) If $\mathbf{a}_{ij}$ and $\mathbf{a}_{kl}$ are leading 1's with $i < k$, then $j < l$.

(iii) If a^i is a non-zero row and a^j is a zero row, then $i < j$.

(iv) If $\mathbf{a}_{ij}$ is a leading 1, then $a_{kj} = 0$ for all $1 \leq k \leq m$ with $k \neq i$.

Observe that Reduced Gauss-Jordan algorithm produced a matrix in reduced row-echelon form.

N2.3 Solving Linear Systems

Theorem N2.3.1. *Let B be the reduced echelon form of the augmented matrix of system of m linear equation in n variables $x_1, \ldots, x_n$. Let $1 \leq d \leq n$. x_j is called a lead variables if Column j of B contains a leading 1 . Otherwise x_j is called a free variable. Let s be the number of lead variables and t the number of free variables. Let $x_{f_1}, \ldots, x_{f_t}$ be the free variables where $f_1 < f_2 < \ldots < f_t$ and put $y_e = x_{f_e}$. Let $x_{l_1}, x_{l_2}, \ldots, x_{l_s}$ be the lead variables where $l_1 < l_2 < \ldots < l_s$.*

Let $1 \leq e \leq t$. If $x_j = x_{l_d}$ is a lead variable, define $b_j = b_{d,n+1}$ and $c_{je} = -b_{d,f_e}$. If x_j is a free variable, define $b_j = 0$ and $c_{je} = 1$, if $x_j = y_e$, and $c_{je} = 0$, if $x_j \neq y_e$.

(a) $n = s + t$, $0 \leq s \leq \min(m, n)$ *and* $\max(n - m, 0) \leq t \leq n$.

(b) Suppose Column $n + 1$ of B contains a leading 1, then the system of the equations has no solutions.

(c) Suppose that Column $n + 1$ of B does not contain a leading 1. Then the solution set of the system of equations is

$$S = \left\{ y_1 \begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \end{pmatrix} + y_2 \begin{pmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{n2} \end{pmatrix} + \ldots + y_t \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{nt} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \middle| \, (y_1, \ldots, y_t) \in \mathbb{R}^t \right\}$$

$$= \{y_1 c_1 + y_2 c_2 + \ldots + y_t c_t \mid (y_1, \ldots, y_t) \in \mathbb{R}^t\}, \textit{ where } c_e = \begin{pmatrix} c_{1e} \\ c_{2e} \\ \vdots \\ c_{ne} \end{pmatrix}$$

Moreover, if $r_1 c_1 + \ldots r_t c_t = s_1 c_1 + \ldots s_t c_t$, then $r_1 = s_1, r_2 = s_2, \ldots, r_t = s_t$.

Proof. (a): Since either variable is either a lead variable or a free variable not both $n = s+t$ and $0 \leq s, t \leq n$. Since each Row of B contains at most one leading 1, $s \leq m$. So $t = n - s \geq n - m$ and (a) holds.

(b): Suppose that the last Column of B contains a leading 1. Say $\mathbf{a}_{i,n+1}$ is a leading 1. Then the first n entries of row i of B are zero, and the equations corresponding to row i is

$$0x_1 + 0x_2 + \ldots + 0 + 0x_n = 1.$$

So $0 = 1$ and the system of equation has no solutions.

(c): Suppose now that the last column of B does nor contain a leading 1. Since B has s leading variables, rows $s+1, s_2 \ldots s_m$ of B are zero rows and can be ignored. Now let $1 \leq d \leq e$ and let $\mathbf{a}_{dj}$ be the leading 1 in Row d. Since the last Column of B does not contain a leading 1, $j \neq n+1$ and so x_j is the d'th leading variable. So $j = l_d$. If x_k is any other leading variable, Condition (iv) of a matrix in reduced echelon form implies that $b_{dk} = 0$. Thus the equation corresponding to Row d of B is

$$x_j + b_{df_1}x_{f_1} + b_{df_1}x_{f_2} + \ldots + b_{df_t}x_{f_t} = b_{d,n+1}$$

and hence equivalent to

$$x_j = -b_{df_1}x_{f_1} - b_{df_1}x_{f_2} - \ldots - b_{df_t}x_{f_t} + b_{d,n+1}$$

Since $y_e = x_{f_e} b_j = b_{d,n+1}$ and $c_{je} = -b_{df_e}$ the linear system of equation is equivalent to

$$(*) \qquad x_j = y_1 c_{j1} + y_2 c_{j+2} + \ldots + y_t c_{jt} + b_j, \quad 1 \leq j \leq n, x_j \text{ is a lead variable}$$

So we obtain a solution by choosing free variables $y_1, y_2, \ldots y_t$ arbitrarily and then compute the leading variables x_j according to (*).

Now lead x_j be free variable. Then $x_j = x_{f_e} = y_e$ for some $1 \leq e \leq t$. Since $c_{jk} = 0$ for $k \neq e$ and $b_j = 0$ We conclude that

$$x_j = y_1 c_{j1} + y_2 c_{j+2} + \ldots + y_t c_{jt} + b_j, \quad 1 \leq j \leq n, x_j \text{ is a lead variable}$$

Together with (*) we conclude that

$$(**) \qquad x_j = y_1 c_{j1} + y_2 c_{j+2} + \ldots + y_t c_{jt} + b_j, \quad 1 \leq j \leq n$$

Writing (**) in vector form we conclude that solution set is

$$\left\{ y_1 \begin{pmatrix} c_{11} \\ c_{21} \\ \vdots \\ c_{n1} \end{pmatrix} + y_2 \begin{pmatrix} c_{12} \\ c_{22} \\ \vdots \\ c_{n2} \end{pmatrix} + \ldots + y_t \begin{pmatrix} c_{1t} \\ c_{2t} \\ \vdots \\ c_{nt} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \,\middle|\, (y_1, \ldots, y_t) \in \mathbb{R}^t \right\}$$

Since $x_{f_e} = y_e$, y_e is uniquely determined by $(x_1, \ldots, x_n)$ and so (c) holds. $\square$

Consider the special case $t = 0$ in part (c) of the previous theorem. Note that $t = 0$ means that none of the variables are free. So all variables are lead variables. So this occurs if the first n Columns of B contain a leading 1, but the last column does not. In this case (c) says that the system of equation has a unique solution namely $x_1 = b_1, x_2 = b_2, \ldots, x_n = b_n$.

Theorem 2.6. *Consider a homogeneous system of m linear equation in n variables. If $n > m$, the system has infinitely many solutions. In particular, it has a non-trivial solutions.*

Proof. Let t be the number of lead variables. By N2.3.1(a), $t \geq n - m > 0$. Since the the last column the augmented matrix is zero we can apply N2.3.1(c). Since $t > 0$, there are infinitely many choices for $(y_1, \ldots, y_t)$ and so the system has infinitely many solutions. $\square$

Chapter N3

Dimension Theory

N7.1 Induction

In the following we will assume the following property of the natural numbers without proof:

Theorem N7.1.1 (Well Ordering Axiom)**.** *Let A be a non-empty set of natural numbers. Then A has minimal element m, that is there exists $m \in A$ with $m \leq a$ for all $m \in \mathbb{N}$.*

Using the Well Ordering Axiom we prove:

Theorem 7.1 (Principal of mathematical induction)**.** *For each $n \in \mathbb{N}$ let S_n be a statement. Suppose that*

(i) S_1 is true.

(ii) If $n \in \mathbb{N}$ and S_n is true, then also S_{n+1} is true.

Then S_n is true for all $n \in \mathbb{N}$.

Proof. Suppose S_a is false for some $a \in \mathbb{N}$. Put $A = \{n \in \mathbb{N} \mid S_n \text{ is false}\}$. Since S_a is false, $a \in A$ and so A is not empty. Thus by the Well Ordering Axiom N7.1.1, A has a minimal element m. So $m \in A$ and $m \leq b$ for all $b \in A$. Since $m \in A$, the definition of A implies that S_m is false. By (i), S_1 is false and so $m \neq 1$. Put $n = m - 1$. Since $m \neq 1$ we have $m \geq 2$ and so $n \geq 1$. Thus n is a positive integers and $n < m$.

By (LR 11), S_n is true or S_n is false. We will show that either case leads to a contradiction.

Suppose that S_n is false. Then $n \in A$ and so since m is minimal element of A, $m \leq n$. a contradiction since $n > m$.

Suppose that S_n is true. Then by (ii) also S_{n+1} is true. But $n + 1 = (m - 1) + 1 = m$ and so S_m is true, a contradiction since $m \in A$ and so S_m is false.

We reach a contradiction to the assumption that S_a is false for some $a \in \mathbb{N}$ and so S_a is true for all $a \in \mathbb{N}$. □

N3.1 Linear Combinations

Definition N3.1.1. *Let* $\mathbf{V}$ *be a vector space,* $n \in \mathbb{N}_0$ *and* $(v_1, \dots v_n)$ *a list of length* n *in* V*. Then, for* $0 \le k \le n$, $\sum_{i=1}^{k} v_i$ *is defined inductively as follows:*

(i) If $k = 0$*, then* $\sum_{i=1}^{k} v_i = \mathbf{0}$*. (So the sum of the empty list is the zero vector).*

(ii) If $k < n$*, then* $\sum_{i=1}^{k+1} v_i = \left(\sum_{i=1}^{k} v_i\right) + v_n$*.*

$\sum_{i=1}^{n} v_i$ *is called the sum of* $(v_1, \dots, v_n)$*. We denote this sum also by*

$$v_1 + \dots + v_n$$

Note that $v_1 + \dots + v_n = \Big(\Big(\dots\big((v_1 + v_2) + v_3\big) + \dots + v_{n-2}\Big) + v_{n-1}\Big) + v_n$. But thanks to the associative and commutative law, this sum is independent of the choice of the parenthesis and also of the order of $v_1, v_2, v_3, \dots, v_n$. A detailed proof of this fact requires a subtle induction argument which we will omit.

Definition 3.1. *Let* $\mathbf{V}$ *be a vector space,* $B = (v_1, \dots, v_n)$ *a list in* V *and* $r = (r_1, r_2, \dots, r_n)$ *a list of real numbers. Then*

$$r_1 v_1 + r_2 v_2 + \dots + r_n v_n$$

is called the linear combination of $(v_1, v_2 \dots, v_n)$ *with respect to the coefficients* $(r_1, \dots, r_n)$

We sometimes denote this linear combination by Br*.*

N3.2 Span

Definition 3.2. *Let* $\mathbf{V}$ *be a vector space.*

(a) Let Z *a subset of* V*. The span of* Z*, denoted by* $\operatorname{span} Z$ *is the set of linear combinations of list in* Z*. So*

$$\operatorname{span} Z = \{r_1 v_1 + r_2 v_2 + \dots + r_n v_n \mid n \in \mathbb{N}_0, (v_1, \dots, v_n) \in Z^n, (r_1, r_2, \dots, r_n) \in \mathbb{R}^n\}$$

(b) Let B *be a list in* V*. The span of* B*, denoted by* $\operatorname{span} B$*, is the set of all linear combinations of* B*. So if* $B = (v_1, \dots, v_n)$*, then*

$$\operatorname{span} B = \{r_1 v_1 + r_2 v_2 + \dots + r_n v_n \mid (r_1, r_2, \dots, r_n) \in \mathbb{R}^n\}$$

(c) Let W *be subset of* V *and* B *a list in* V *or a subset of* V*. We say that* B *spans* W *(or that* B *is a spanning list (set) of* W *) if* $W = \operatorname{span} B$

Since $\mathbf{0}$ is the only linear combination of $()$, $\operatorname{span}() = \operatorname{span} \emptyset = \{\mathbf{0}\}$.

Lemma N3.2.2. *Let V be a vector space and $(v_1, \ldots, v_n)$ a list in V. Then*

$$\operatorname{span}(v_1, \ldots, v_n) = \operatorname{span}\{v_1, \ldots, v_n\}.$$

Proof. Put $Z = \{v_1, \ldots, v_n\}$. Then $(v_1, \ldots, v_n)$ is a list in Z and so $\operatorname{span}(v_1, \ldots, v_n) \subseteq \operatorname{span} Z$.

Let $u \in \operatorname{span} Z$. Then there exist a list $(w_1, \ldots, w_m)$ in Z and $(r_1, \ldots, r_m)$ in $\mathbb{R}$ with

$$(*) \qquad u = r_1 w_1 + \ldots + r_m w_m.$$

We claim that u is a linear combination of $(v_1, \ldots, v_n)$.

If $m = 0$, then $u = \mathbf{0} = 0v_1 + \ldots + 0v_n$ and the claim holds.

Suppose the claim holds for $m - 1$. Then

$$(**) \qquad r_1 w_1 + \ldots + r_{m-1} w_{m-1} = s_1 v_1 + \ldots + s_n v_n$$

for some $(s_1, \ldots, s_n) \in \mathbb{R}^n$. Since $w_m \in Z$, $w_m = v_i$ for some $1 \leq i \leq n$ and so

$$(* * *) \qquad r_m w_m = t_1 v_1 + \ldots + t_n v_n$$

where $t_i = r_m$ and $t_j = 0$ for $1 \leq j \leq n$ with $j \neq 1$. We compute

$$\begin{array}{rcll} u & = & r_1 w_1 + \ldots + r_m w_m & - (*) \\ & = & (r_1 w_1 + \ldots + r_{m-1} w_{m-1}) + r_m w_m & - \text{definition of } ' + \ldots +' \\ & = & (s_1 v_1 + \ldots + s_n v_n) + r_m w_m & - (**) \\ & = & (s_1 v_1 + \ldots + s_n v_n) + (t_1 v_1 + \ldots + t_n v_n) & - (***) \\ & = & (s_1 + t_1) v_1 + \ldots + (s_n + t_n) v_n & - \text{D.2.1} \end{array}$$

So the claim holds for $m + 1$ and hence by the principal of induction for all non-negative integers n.

Thus u is a linear combination of $(v_1, \ldots, v_n)$ and so $\operatorname{span} Z \subseteq \operatorname{span}(v_1, \ldots, v_n)$. We already proved the reverse inclusion and so the $\operatorname{span} Z = \operatorname{span}(v_1, \ldots, v_n)$. □

Theorem 3.3. *Let $\mathbf{V}$ be a vector space and Z a subset of V.*

(a) $\operatorname{span} Z$ *is a subspace of* $\mathbf{V}$ *and* $Z \subseteq \operatorname{span} Z$.

(b) Let $\mathbf{W}$ *be a subspace of* $\mathbf{V}$ *with* $Z \subseteq W$. *Then* $\operatorname{span} Z \subseteq W$.

Proof. (a) Since $()$ is a list in Z and 0_V is a linear combination of $()$, $0_V \in \operatorname{span} Z$.

Let $x, y \in \operatorname{span} Z$ and $a \in \mathbb{R}$. Then $x = r_1 v_1 + r_2 v_2 + \ldots + r_n v_n$ and $y = s_1 w_1 + s_2 w_2 + \ldots + s_m w_m$ for some lists $(r_1, \ldots, r_n)$ and $(s_1, \ldots, s_m)$ in $\mathbb{R}$ and $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m) \in Z$. Thus

$$x + y = r_1v_1 + r_2v_2 + \ldots + r_nv_n + s_1w_1 + s_2w_2 + \ldots + s_mw_m$$

is a linear combination of the list $(v_1, \ldots, v_n, w_1, \ldots, w_m)$ in Z. Thus $x + y \in \operatorname{span} Z$ and so $\operatorname{span} Z$ is closed under addition.

Also

$$ax = (ar_1)v_1 + (ar_2)v_2 + \ldots + (ar_n)v_n$$

and so $ax \in \operatorname{span} Z$. We verified the three conditions of the Subspace Theorem and so $\operatorname{span}(Z)$ is a subspace.

If $z \in Z$, then $z = 1z$ is a linear combination of the list (z) in Z and so $Z \subseteq \operatorname{span} Z$.

(b) This follows easily from the fact that W is closed under addition and scalar multiplication, but we will give a detailed induction proof. Let B be a list of length n in Z. We will show by induction on the n that any linear combination of B is contained in W.

Suppose first that $n = 0$. Then $\mathbf{0}$ is the only linear combination of B. Also by the Subspace Theorem $\mathbf{0} \in W$. So indeed every linear combination of B is in W.

Suppose any linear combination of a list of length n in Z is contained in W and let $(v_1, \ldots, v_n, v_{n+1})$ be a list of length $n+1$ in Z. sLet $(r_1, \ldots, r_n, r_{n+1}) \in \mathbb{R}^{n+1}$. Then by the definition of 'span', $\sum_{i=1}^{n} r_iv_i \in \operatorname{span}(v_1, v_2, \ldots, v_n)$. By the induction assumption,

$$\sum_{i=1}^{n} r_iv_i \in W. \tag{1}$$

By the Subspace Theorem, W is closed under scalar multiplication and since $v_{n+1} \in Z \subseteq W$, we conclude

$$r_{n+1}v_{n+1} \in W. \tag{2}$$

By the definition of 'sum'

$$\sum_{i=1}^{n+1} r_iv_i = \left(\sum_{i=1}^{n} r_iv_i\right) + r_{n+1}v_{n+1} \tag{3}$$

By the Subspace Theorem, W is closed under addition and so (1),(2) and (3) show that $\sum_{i=1}^{n+1} r_iv_i \in W$.

We proved that all linear combinations of $(v_1, \ldots, v_{n+1})$ are in W so by the Principal of Induction any linear combination of any list in Z is contained in W. So by the definition of span, $\operatorname{span} Z \subseteq W$. □

In less precise terms the preceding theorem means that $\operatorname{span} Z$ is the smallest subspace of V containing Z.

Corollary N3.2.4. *Let $\mathbf{V}$ be a vector space, W a subspace of V and Z a spanning set for W.*

(a) If U is a subspace of W with $Z \subseteq U$, then $U = W$.

(b) Let $X \subseteq W$ and suppose that each z in Z is a linear combination of a list in X, then X spans W.

Proof. (a) Since Z is a spanning set for W, $W = \operatorname{span} Z$. Since $Z \subseteq U$ 3.3(b) gives $W \subseteq U$. By assumption $U \subseteq W$ and so $U = W$.

(b) Put $U = \operatorname{span} X$. By 3.3(a), U is a subspace of W. By assumption each $z \in Z$ is linear combination of a list in X and so $Z \subseteq U$. Thus by (a), $U = W$ and so $\operatorname{span} X = W$ and X spans W. □

N3.3 Linear Independence

Definition 3.4. *Let $\mathbf{V}$ be a vector space and $(v_1, \ldots, v_n)$ a list of vectors in V. We say that $(v_1, \ldots, v_n)$ is linearly independent if for all $(r_1, r_2, \ldots r_n) \in \mathbb{R}^n$,*

$$r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = \mathbf{0} \quad \Longrightarrow \quad r_1 = 0, r_2 = 0, \ldots, r_n = 0.$$

Lemma N3.3.2. *Let $\mathbf{V}$ be a vector space, $(v_1, \ldots, v_n)$ a list in V and $i \in \mathbb{N}$ with $1 \leq i \leq n$. Then the following three statements are equivalent.*

(a) There exists $(r_1, r_2 \ldots, r_n) \in \mathbb{R}^n$ such that

$$r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = \mathbf{0}$$

and $r_i \neq 0$.

(b) $v_i \in \operatorname{span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$

(c) $\operatorname{span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_n) = \operatorname{span}(v_1, \ldots, v_n)$

Proof. Assume without loss that $i = 1$.

(a) $\Longrightarrow$ (b): Suppose that (a) holds. Then $r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = \mathbf{0}$. Since $r_1 \neq 0$ we can solve for v_1 and get

$$v_1 = (-\frac{r_2}{r_1})v_2 + (-\frac{r_3}{r_1})v_3 + \ldots + (-\frac{r_n}{r_1})v_n$$

and so by the definition of span,

$$v_1 \in \operatorname{span}(v_2, v_3, \ldots, v_n)$$

Thus (a) holds.

(b) $\Longrightarrow$ (c): Suppose (b) holds. Then $v_1 \in \operatorname{span}(v_2, \ldots, v_n)$. By 3.3(a), $v_j \in \operatorname{span}(v_2, v_3, \ldots, v_n)$ for all $2 \leq j \leq n$. Hence $v_i \in \operatorname{Span}(v_2, \ldots, v_n)$ for all $1 \leq i \leq n$. Thus by N3.2.4, $\operatorname{span}(v_2, \ldots, v_n) = \operatorname{span}(v_1, \ldots, v_n)$.

(c) $\Longrightarrow$ (a): Suppose (c) hold, that is $\operatorname{span}(v_2, \ldots, v_n) = \operatorname{span}(v_1, \ldots, v_n)$. Then $v_1 \in \operatorname{span}(v_2, \ldots, v_n)$ and so

$$v_1 = r_2 v_2 + \ldots + r_n v_n$$

for some $r_2, \ldots, r_n \in \mathbb{R}$. Thus

$$(-1)v_1 + r_2 v_2 + \ldots + r_n v_n = \mathbf{0}.$$

Put $r_1 = -1$. Then $r_1 \neq 0$ and $r_1 v_1 + r_2 v_2 \ldots + r_n v_n = \mathbf{0}$. Therefore (a) holds.

$\square$

Theorem 3.5. *Let* $\mathbf{V}$ *be a vector space and* $(v_1, \ldots, v_n)$ *be list of vectors in* V*. Then the following are equivalent:*

(a) $(v_1, \ldots, v_n)$ *is linearly independent.*

(b) For each $v \in V$ *there exists at most one* $(r_1, \ldots, r_n) \in \mathbb{R}^n$ *with*

$$v = r_1 v_1 + r_2 v_2 + \ldots + r_n v_n.$$

(c) For all $1 \leq i \leq n$*,*

$$v_i \notin \operatorname{span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n).$$

(d) For all $1 \leq i \leq n$*,*

$$v_i \notin \operatorname{span}(v_1, \ldots, v_{i-1}).$$

(e) There exists $0 \leq k < n$ *such that* $(v_1, \ldots, v_k)$ *is linearly independent and,*

$$v_i \notin \operatorname{span}(v_1, \ldots, v_{i-1})$$

for all $k+1 \leq i \leq n$*,*

(f) $(v_1, \ldots, v_{n-1})$ *is linearly independent and* $v_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$*.*

Proof. (a) $\Longrightarrow$ (b): Suppose $(v_1, \ldots, v_n)$ is linearly independent. Let $v \in V$ and suppose there exist $(r_1, \ldots, r_n)$ and $(s_1, \ldots, s_n) \in \mathbb{R}^n$ with $v = r_1 v_1 + r_2 v_2 + \ldots + r_n v_n$ and $v = s_1 v_1 + s_2 v_2 + \ldots + s_n v_n$. Then

$$r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = s_1 v_1 + s_2 v_2 + \ldots + s_n v_n,$$

and so

$$(r_1 - s_1)v_1 + (r_2 - s_2)v_2 + \ldots + (r_n - s_n)v_n = \mathbf{0}.$$

Since $(v_1, \ldots, v_n)$ is linear independently this means $r_1 - s_1 = r_2 - s_2 = \ldots = r_n - s_n = 0$ and so $r_1 = s_1, r_2 = s_2, \ldots r_n = s_n$. Thus (b) holds.

(b) $\Longrightarrow$ (c): Suppose (b) holds. We will show that (c) holds via a contradiction proof. So assume that (c) is false. Then there exists $1 \leq i \leq n$ with

$$v_i \in \operatorname{span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots v_n)$$

To simplify notation we assume (without loss) that $i = 1$ By definition of span

$$v_1 = r_2 v_2 + r_2 v_3 + \ldots + r_n v_n$$

for some $r_2, \ldots, r_n \in \mathbb{R}$. Thus

$$1v_1 + 0v_2 + 0v_3 + \ldots + 0v_n = 0v_1 + r_2 v_2 + r_3 v_3 + \ldots + r_n v_n.$$

(b) shows that $1 = 0, 0 = r_2, \ldots, 0 = r_n$, a contradiction.

(c) $\Longrightarrow$ (d): Since $\operatorname{span}(v_1, \ldots, v_{i-1}) \subseteq \operatorname{span}(v_1, \ldots, v_{i-1}, v_{i+1}, \ldots, v_n)$, this implication is obvious.

(d) $\Longrightarrow$ (e): If (d) hold, then (e) holds with $k = 0$.

(f) $\Longrightarrow$ (a): Suppose (f) holds and let $(r_1, r_2 \ldots, r_n) \in \mathbb{R}$ with

$$r_1 v_1 + r_2 v_2 + \ldots + r_n v_n = \mathbf{0}$$

Since $v_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$, N3.3.2 shows that $r_n = 0$. Thus

$$r_1 v_1 + r_2 v_2 + \ldots + r_{n-1} v_{n-1} = \mathbf{0}$$

Since $(v_1, \ldots, v_{n-1})$ is linearly independent this implies $r_1 = r_2 = \ldots = r_{n-1} = 0$. Since also $r_n = 0$, $(v_1, \ldots, v_n)$ is linearly dependent.

(e) $\Longrightarrow$ (f): Suppose (d) holds. Let $k \leq i \leq n$. We will show by induction that $(v_1, \ldots, v_i)$ is linearly independent. For $i = k$, this holds by assumption. Suppose now that $k \leq i < n$ and $(v_1, \ldots, v_i)$ is linearly independent. By assumption, $v_{i+1} \notin \operatorname{span}(v_1, \ldots, v_i)$ and so the already proven implication '(f) $\Longrightarrow$ (a)' applied with $n = i + 1$ shows that $(v_1, \ldots, v_{i+1})$ is linearly independent. Thus by the principal of induction, $(v_1, \ldots, v_{n-1})$ is linearly independent. Also by assumption $v_n \notin \operatorname{span}(v_1, \ldots, v_{n-1})$ and so (f) holds.

$\square$

N3.6 Coordinates

Definition 3.6. *Let* $\mathbf{V}$ *be a vector space. A basis for* $\mathbf{V}$ *is a linearly independent spanning list for* V.

Theorem 3.17. *Let* $\mathbf{V}$ *be a vector space and* $(v_1, \ldots, v_n)$ *be a list in* V. *Then* $(v_1, v_2, \ldots, v_n)$ *is a basis for* V *if and only of for each* $v \in V$ *there exists a unique*

$$(r_1, r_2, \ldots, r_n) \in \mathbb{R}^n \text{ with } v = r_1v_1 + r_2v_2 + \ldots + r_nv_n$$

Proof. By definition $(v_1, \ldots, v_n)$ is basis if and only if its spans V and is linearly independent. By definition its spans V if and only if for each $v \in V$ there exists a $(r_1, r_2, \ldots, r_n) \in \mathbb{R}^n$ with $v = r_1v_1 + r_2v_2 + \ldots + r_nv_n$. And by 3.5 it is linearly independent if and only if for each v in V there exists a most one $(r_1, r_2, \ldots, r_n) \in \mathbb{R}^n$ with $v = r_1v_1 + r_2v_2 + \ldots + r_nv_n$. □

Definition 3.16. *Let* $\mathbf{V}$ *be a vector space with a basis* $B = (v_1, \ldots, v_n)$. *For* $v \in V$ *let* $(r_1, r_2, \ldots, r_n)$ *be the unique list in* $\mathbb{R}$ *with* $v = r_1v_1 + r_2v_2 + \ldots + r_nv_n$. *Then* $(r_1, r_2, \ldots, r_n)$ *is called the coordinate vector of* v *with respect to* B *and is denoted by* $[v]_B$. *The function* $C_B : V \to \mathbb{R}^n$ *defined by* $C_B(v) = [v]_B$ *is called the coordinate function of* V *with respect to* B.

Example N3.6.4. *Let* $E = (e_1, \ldots, e_n)$ *be the standard basis for* $\mathbb{R}^n$. *Then* $[x]_E = x$ *for all* $x \in \mathbb{R}^n$ *and so* $C_E = \mathrm{id}_{\mathbb{R}^n}$.

Proof. Let $x = (x_1, \ldots, x_n)$. Then

$$\begin{aligned}
x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} &= \begin{pmatrix} x_1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ x_2 \\ \vdots \\ 0 \end{pmatrix} + \ldots + \begin{pmatrix} 0 \\ 0 \\ \vdots \\ x_n \end{pmatrix} \\
&= x_1 \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix} + x_2 \begin{pmatrix} 0 \\ 1 \\ \vdots \\ 0 \end{pmatrix} + \ldots + x_n \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix} \\
&= x_1e_1 + x_2e_2 + \ldots + x_ne_n
\end{aligned}$$

and so

$$[x]_E = (x_1, \ldots, x_n) = x$$

It follows that $C_E(x) = x$ and so $C_E = \mathrm{id}_{\mathbb{R}^n}$. □

N3.4 Bases

Theorem N3.4.1. *Let* $\mathbf{V}$ *be a vector space,* $(v_1, \ldots, v_n)$ *a linearly independent list in* V *and* Z *a subset of* V*. Put* $W = \operatorname{span}\big(\{v_1, \ldots, v_n\} \cup Z\big)$ *and let* $(u_1, \ldots, u_l)$ *be a list in* Z*. Then the following three statements are equivalent*

(a) $(v_1, \ldots, v_n, u_1, \ldots, u_l)$ *is a basis for* W*.*

(b) $(u_1, u_2, \ldots, u_l)$ *is a list in* Z *minimal such that* $(v_1, \ldots, v_n, u_1, \ldots, u_l)$ *spans* W*.*
That is

$$(v_1, \ldots, v_n, u_1, \ldots, u_l)$$

spans W*, but for all* $1 \leq i \leq l$*,*

$$(v_1, \ldots, v_n, u_1, \ldots, u_{i-1}, u_{i+1} \ldots, u_l)$$

does not span W

(c) $(u_1, u_2, \ldots, u_l)$ *is a list in* Z *maximal such that* $(v_1, \ldots, v_n, u_1, \ldots, u_l)$ *is linearly independent.*
That is

$$(v_1, \ldots, v_n, u_1, \ldots, u_l)$$

is linearly independent but for all $z \in Z$*,*

$$(v_1, \ldots, v_n, u_1, \ldots, u_l, z)$$

is linearly dependent.

Proof. (a) $\Longrightarrow$ (b): Let $1 \leq i \leq n$. Since $(v_1, v_2, \ldots, v_n, u_1, \ldots u_l)$ is linear independent, 3.5 shows that $v_i \notin \operatorname{span}(v_1, v_2, \ldots, v_n, u_1, \ldots u_{i-1}, \ldots, u_{i+1} \ldots, u_l)$. So

$$(v_1, v_2, \ldots, v_n, u_1, \ldots u_{i-1}, u_{i+1} \ldots, u_l).$$

does not span W. Since $(v_1, v_2, \ldots, v_n, u_1, \ldots u_l)$ is a basis of W it spans W and (b) holds.

(b) $\Longrightarrow$ (a): Let $1 \leq i \leq l$. By minimality of $(u_1, \ldots, u_l)$.

$$\operatorname{span}(v_1, \ldots, v_n, u_1, \ldots u_{i-1}, u_{i+1}, \ldots, u_l) \neq W = \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots, u_l)$$

and so by by N3.3.2,

$$u_i \notin \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots, u_{i-1}, u_{i+1}, \ldots, u_l)$$

and so also

$$u_i \notin \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots, u_{i-1})$$

3.5(e) now shows that $(v_1, \ldots, v_n, u_1, \ldots, u_n)$ is linearly independent. It also spans W and so is a basis for W.

(a) $\Longrightarrow$ (c): Let $z \in Z$. Then $z \in W = \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots, u_l)$ and so by 3.5, $(v_1, \ldots, v_n, u_1, \ldots u_l, z)$ is linearly dependent. Thus (c) holds.

(c) $\Longrightarrow$ (a): Put $U = \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots u_l)$ and let $z \in Z$. We will show that $z \in U$. By the maximality of $(u_1, \ldots, u_l)$,

$$(v_1, \ldots, v_n, u_1, \ldots u_l, z)$$

is linearly dependent. Since $(v_1, \ldots, v_n, u_1, \ldots u_l)$ is linearly independent we conclude from 3.5(f) that

$$z \in \operatorname{span}(v_1, \ldots, v_n, u_1, \ldots u_l) = U$$

Thus $Z \subseteq U$. . Since also $v_i \in U$ for all $1 \leq i \leq n$, N3.2.4 shows that $U = W$. Thus $(v_1, v_2, \ldots, v_n, u_1, \ldots u_l)$ spans W. It also linearly independent, so its a basis for W. Hence (a) holds.

□

Definition 3.8. *A vector space is called finite dimensional if it has a finite spanning set. Otherwise, it is called infinite dimensional. If V is finite dimensional, the minimal size of a spanning set is called the dimension of V and is denoted by* $\dim V$*.*

Theorem 3.11 (Contraction Theorem)**.** *Any spanning list of a vector space has sublist which is basis.*

Proof. Let $(w_1, w_2, \ldots, w_m)$ be a spanning list and $(u_1, \ldots, u_l)$ a sublist minimal such that $(u_1, \ldots, u_l)$ spans W. Then by N3.4.1 applied with $n = 0$, $(u_1, \ldots, u_l)$ is basis. □

Corollary N3.4.4. *Let* $\mathbf{V}$ *be a finite dimensional vector space and put* $n = \dim \mathbf{V}$*. Then every spanning list of V of length n is a basis for V. In particular, V has a basis of length n.*

Proof. By definition of $n = \dim V$, V has a spanning list D of length n and every spanning list as length at least n.

Now let D be any spanning list of length n. By 3.11, D has a sublist B which is basis. Then B spans V and so B has length at least n. Thus $D = B$ and so D is a basis. □

We will see later (3.10) that all bases of a finite dimensional vector space have length $\dim V$.

Theorem 3.13 (Expansion Theorem)**.** *Any linearly independent list in a finite-dimensional vectors space is the sublist of a basis.*

Proof. Let $(v_1, \ldots, v_n)$ be a linearly independent list in the vector space V. Since V is finite dimensional, V has a spanning list $(w_1, w_2, \ldots, w_m)$. Thus $V = \operatorname{span}(v_1, \ldots, v_n, w_1, \ldots, w_m)$. Let $(u_1, \ldots, u_l)$ be a sublist of $(w_1, \ldots, w_m)$ maximal such that $(v_1, \ldots, v_n, u_1, \ldots, u_l)$ is linearly independent. Then by N3.4.1 $(v_1, \ldots, u_1, \ldots, u_l)$ is a basis for V. $\square$

N3.7 Col, Row and Nul

Definition N3.7.1. *Let A be an $m \times n$-matrix and $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then*

$$Ax = x_1a_1 + x_2a_2 + \ldots + x_na_n.$$

So $Ax = Bx$ where B is the list of columns of A.

Let $x, y \in \mathbb{R}^n$. Recall that according to 3.1 xy is defined as $y_1x_1 + y_2x_2 + \ldots + y_nx_n$. Note that $xy = yx$. Also entry i of Ax is

$$x_1a_{i1} + x_2a_{i2} + \ldots + x_na_{in} = a^ix = xa^i = a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n.$$

Definition 6.21,6.23. *Let n and m be positive integer and A an $m \times n$-matrix. Then*

(a) $\operatorname{Col}A = \operatorname{span}(a_1, \ldots, a_n)$.

(b) $\operatorname{Row}A = \operatorname{span}(a^1, \ldots, a^m)$.

(c) $\operatorname{Nul}A = \{x \in \mathbb{R}^n \mid Ax = \mathbf{0}\}$

Note here that Nul$A$ consists of the solutions of the homogeneous system of linear equation

$$x_1a_1 + x_2a_2 + \ldots + x_na_n = \mathbf{0}$$

Lemma 6.24. *Let A be matrix and B a matrix obtained from A be sequence of elementary row operations.*

(a) $\operatorname{Nul}A = \operatorname{Nul}B$

(b) $\operatorname{Row}A = \operatorname{Row}B$.

Proof. (a) holds since row operations do not change the solutions of an homogeneous system of linear equations.

(b) A simple induction argument shows that we may assume that B is obtained from A by just one elementary row operation. We will first show that $\operatorname{Row}B \subseteq \operatorname{Row}A$.

By definition of an elementary row operation, any row of B is either a row of A, a non-zero scalar multiple of a row of A or the sum of a row of A and a scalar multiple of a row of A. So any row of B is linear combination of rows of A.

In particular, all rows of A are contained in $\operatorname{Row}A = \operatorname{span}(a^1, \ldots, a^m)$. Thus by 3.3, $\operatorname{Row}B = \operatorname{span}(b^1, \ldots, b^m) \subseteq \operatorname{Row}A$.

Since the inverse of a elementary row operation is also an elementary row operation, we conclude that also $\mathrm{Row}A \subseteq \mathrm{Row}B$. So $\mathrm{Row}A = \mathrm{Row}B$ and (a) holds.

□

Lemma N3.7.4. *Let A be an $m \times n$- matrix and B a matrix obtained from A by sequence of elementary row operations. Let l be positive integer and $1 \leq j_1 < j_2 < \ldots < j_l \leq n$. Put $c_k = a_{j_k}$ and $d_k = b_{j_k}$.*

(a) Let $(x_1, \ldots, x_l)$ and $(y_1, \ldots, y_n)$ be list in $\mathbb{R}$. Then

$$x_1c_1 + \ldots + x_lc_l = y_1a_1 + \ldots + y_na_n \iff x_1d_1 + \ldots + x_ld_l = y_1b_1 + \ldots + y_nb_n$$

(b) $(c_1, \ldots, c_l)$ is basis for $\mathrm{Col}A$ if and only if $(d_1, \ldots, d_l)$ is a basis for $\mathrm{Col}B$.

Proof. (a) For $1 \leq i \leq n$, put $z_i = x_{j_k}$ if $i = j_k$ for some $1 \leq k \leq l$ and $z_i = 0$ otherwise. Then

$$\begin{aligned}
& x_1c_1 + \ldots + x_lc_l = y_1a_1 + \ldots + y_na_n \\
\iff\ & z_1a_1 + \ldots + z_na_n = y_1a_1 + \ldots + y_na_n \\
\iff\ & (z_1 - y_1)a_1 + \ldots + (z_n - y_n)a_n = \mathbf{0} \\
\iff\ & (z_1 - y_1, \ldots, z_n - y_n) \in \mathrm{Nul}A \\
\iff\ & (z_1 - y_1, \ldots, z_n - y_n) \in \mathrm{Nul}B \\
\iff\ & x_1d_1 + \ldots + x_ld_l = y_1b_1 + \ldots + y_nb_n
\end{aligned}$$

So (a) holds.

(b) By 3.17 $(c_1, c_2, \ldots, c_l)$ is a basis for $\mathrm{Col}A$ if and only if for each $v \in \mathrm{Col}A$ there exists a unique $x = (x_1, \ldots x_l) \in \mathbb{R}^l$ with $x_1c_1 + \ldots + x_lc_l = v$. By definitions of $\mathrm{Col}A$, $(a_1, \ldots, a_n)$ spans $\mathrm{Col}A$. So $v = y_1a_1 + \ldots + y_na_n$ for some $(y_1, \ldots, y_n) \in \mathbb{R}^n$. Thus $x_1c_1 + \ldots + x_lc_l = v$ if and only if

$$x_1c_1 + \ldots + x_lc_l = y_1a_1 + \ldots + y_na_n$$

Hence

1°. *$(c_1, \ldots, c_l)$ is a basis for* Col*A if and only if for each $(y_1, \ldots, y_n) \in \mathbb{R}^n$ there exists a unique $(x_1, \ldots, x_l) \in \mathbb{R}^l$ with $x_1c_1 + \ldots + x_lc_l = y_1a_1 + \ldots + y_na_n$.*

The same argument shows

2°. *$(d_1, \ldots, d_l)$ is a basis for* Col*B if and only if for each $(y_1, \ldots, y_n) \in \mathbb{R}^n$ there exists a unique $(x_1, \ldots, x_l) \in \mathbb{R}^l$ with $x_1d_1 + \ldots + x_ld_l = y_1b_1 + \ldots + y_nb_n$*

(a) now shows that (b) holds. □

Theorem N3.7.5. *Let A be a $m \times n$-matrix and B its reduced echelon form. Then:*

(a) *Let s be the number of lead variables of A. Let x_{l_k} be the k'th lead variables of A. Then $b_{l_k} = e_k$ for all $1 \leq k \leq s$, where $(e_1, \ldots, e_m)$ is standard basis for $\mathbb{R}^m$.*

(b) *The non-zero rows of B form a basis of* Row$A$.

(c) *The columns of A corresponding to the lead variables of B form a basis for* Col$A$.

(d) *Let t the number of free variables of A. Let $c_1, \ldots, c_t$ be the vectors in $\mathbb{R}^n$ defined in N2.3.1. Then $(c_1, \ldots, c_t)$ is a basis for* Nul$A$.

Proof. Let $1 \leq k \leq s$. Observe that by definition of the reduced echelon form, the leading 1 in row k of B is the only non-zero entry in Column l_k of B. Thus

1°. $b_{l_k} = e_k$ *for all* $1 \leq k \leq s$, *that is* $b_{il_k} = 1$ *if* $i = k$ *and* $b_{il_k} = 0$, *if* $i \neq 0$.

In particular, (a) holds.

Note $(b^1, \ldots, b^s)$ is the list of non-zero rows of B. Suppose that $\sum_{i=1}^s r_i b^i = \mathbf{0}$. From (1°) we see that that the l_k entry of $\sum_{i=1}^s r_i b^i$ is r_k. So $r_k = 0$ for all $1 \leq k \leq s$ and $(b^1, \ldots, b^s)$ is linearly independent. $(b^1, \ldots, b^s)$ also spans Row$B$ and so $(b^1, \ldots, b^s)$ is a basis for Row$B$. By 6.24 Row$A$ = Row$B$ and so (b) holds.

(c) Note that $b_{ij} = 0$ for all $s < i \leq m$ and $1 \leq j \leq n$. So if $r = (r_i)_{i=1}^m \in$ Col$B$, then $r_i = 0$ for all $s < i \leq m$. Thus there exists a unique $(u_1, \ldots, u_s) \in \mathbb{R}^s$ with $r = \sum_{i=1}^s u_i e_i$, namely, $u_i = r_i$ for all $1 \leq i \leq s$. Thus $(e_1, \ldots, e_s)$ is a basis for Col$B$. From (1°) we conclude that $(b_{l_1}, \ldots, b_{l_s})$ is a basis for Col(B). Hence by N3.7.4, $(a_{l_1}, \ldots, a_{l_s})$ is a basis for Col(A).

(d) Note that Nul$A$ is the set of solutions of the linear system of equations $x_1 a_1 + \ldots, x_n a_n = \mathbf{0}$. By N2.3.1 each solution can be uniquely written as a linear combination of $(c_1, \ldots, c_t)$. So by 3.17 $c_1, \ldots, c_t$ is basis for Nul$A$. □

N3.5 Dimension

Lemma N3.5.1. *Let* $\mathbf{V}$ *be vector space and $(v_1, \ldots, v_n)$ a spanning list for V. Let $w \in V$ and $1 \leq i \leq n$ and suppose that $w \notin$* span$(v_1, \ldots v_{i-1}, v_{i+1}, \ldots, v_n)$. *Then*

$$(v_1, \ldots, v_{i-1}, w, v_{i+1}, \ldots, v_n)$$

spans V.

Proof. Without loss $i = 1$. Since $(v_1, v_2, \ldots, v_n)$ spans V, $w = r_1 v_1 + \ldots r_n v_n$ for some $r_1, r_2 \ldots, r_n \in V$. Since $w \notin \operatorname{span}(v_2, v_3, \ldots, v_n)$ we have $r_1 \neq 0$. Observe that

$$(-1)w + r_1 v_1 + \ldots r_n v_n = \mathbf{0}.$$

Thus by N3.3.2,

$$\operatorname{span}(w, v_2, \ldots, v_n) = \operatorname{span}(w, v_1, v_2, \ldots, v_n) = V.$$

□

Theorem N3.5.2. *Let* $\mathbf{V}$ *be vector space and let* $(v_1, \ldots, v_n)$ *and* $(w_1, \ldots, w_m)$ *be spanning lists for* V*. Then there exists a sublist* $(u_1, u_2, \ldots, u_l)$ *of* $(w_1, \ldots, w_m)$ *of length at most* n *which spans* V*.*

Proof. For $0 \leq k \leq n$ let S_k be the following statement:

(S_k): There exists a sublist $(u_1, \ldots, u_l)$ of $(w_1, \ldots, w_m)$ of length at most k such that $(u_1, u_2, \ldots, u_l, v_{k+1}, v_{k+2}, \ldots, v_n)$ spans V.

Note that S_n is the statement we would like to prove. We will use induction to show that S_k holds for all $0 \leq k \leq n$.

Since $(v_1, v_2, \ldots, v_n)$ is a basis for V its spans V. So S_0 holds with $(u_1, \ldots, u_l) = ()$, the empty list.

Suppose now that S_k hold and $k < n$. So there a sublist $(u_1, \ldots, u_l)$ of $(w_1, \ldots, w_m)$ of length at most k such that

$$(*) \qquad (u_1, u_2, \ldots, u_l, v_{k+1}, v_{k+2}, \ldots, v_n) \text{ spans } V.$$

If $(u_1, u_2, \ldots, u_l, v_{k+2}, \ldots, v_n)$ spans V, then S_{k+1} holds with the sublist $(u_1, \ldots, u_l)$ of $(w_1, \ldots, w_m)$.

Hence we may assume that $(u_1, u_2, \ldots, u_l, v_{k+2}, \ldots, v_n)$ does not span V. Since $(w_1, \ldots, w_m)$ spans V we conclude from N3.2.4 that

$$(**) \qquad w_i \notin \operatorname{Span}(u_1, u_2, \ldots, u_l, v_{k+2}, \ldots, v_n)$$

for some $1 \leq i \leq m$. Thus (*), (**) and N3.5.1 imply that $(u_1, \ldots, u_l, w_i, v_{k+2}, \ldots v_n)$ is a spanning set of V. By Theorem 3.3, $u_j \in \operatorname{Span}(u_1, u_2, \ldots, u_l, v_{k+2}, \ldots, v_n)$ and so by (**) $w_i \neq u_j$ for all $1 \leq i \leq l$. Thus $(u_1, \ldots, u_l, w_i)$ is (possible after reordering) a sublist of $(w_1, w_2 \ldots, w_n)$ of length $l + 1$. Since $l \leq k$, $l + 1 \leq k + 1$ and so S_{k+1} holds.

We proved that S_k implies S_{k+1} and so by the Principal of induction, S_n holds. □

Theorem 3.10. *Let* $\mathbf{V}$ *be a finite dimensional vector space. Then all bases of* V *have length* $\dim V$*.*

Proof. Let $(w_1, \ldots, w_m)$ be a basis for V and $n = \dim V$. Then by definition of 'dimension' there exists a spanning list $(v_1, \ldots, v_n)$ of length n. Since $(v_1, \ldots, v_n)$ and $(w_1, \ldots, w_m)$ span V we conclude from N3.5.2 there exists a sublist $(u_1 \ldots, u_l)$ of $(w_1, \ldots, w_m)$ of length at most n which spans V. Since $(w_1, \ldots, w_m)$ is a basis for V N3.4.1 implies that $(w_1, \ldots, w_m)$ is minimal spanning sublist of itself. So $m = l$ and thus $m \leq n$. Since by definition $n = \dim V$ is the minimal length of a spanning list, $m \geq n$ and therefore $m = l$. □

Theorem 3.9 (Comparison Theorem)**.** *Let* V *be a finite dimensional vector space,* U *a linear independent list in* V*,* B *a basis for* V *and* S *a spanning list for* V*. Then*

$$\operatorname{length} U \leq \dim V = \operatorname{length} B \leq \operatorname{length} S$$

Proof. By 3.13 U is contained in a basis B' of B. Then $\operatorname{length} U \leq \operatorname{length} B'$. By 3.10 $\operatorname{length} B' = \dim V = \operatorname{length} B$ and by definition of $\dim V$, $\dim V \leq \operatorname{length} S$. So

$$\operatorname{length} U \leq \operatorname{length} B' = \operatorname{length} B = \dim V \leq \operatorname{lengthS}$$

□

Corollary N3.5.5. *Let V be an n-dimensional vector space and B a list of length n in V. Then the following are equivalent*

(a) B is basis for V.

(b) B is linearly independent.

(c) B spans V.

Proof. (a) $\Longrightarrow$ (b): By definition any basis is linearly independent and so (a) implies (b).

(b) $\Longrightarrow$ (c): Suppose B is linearly independent. Then by the Expansion Theorem 3.13, B is a sublist of a basis D. By 3.10 D has length n and since B has also length n, $B = D$.

(c) $\Longrightarrow$ (a): Suppose B spans $\mathbf{V}$. Then by the Contraction Theorem 3.11, B has sublist D which is basis. By 3.10 D has length n and since B has also length n, $B = D$. Thus B is a basis and so B spans V. □

Theorem 6.25. *Let A be an $m \times n$-matrix. Then*

(a) $\dim \operatorname{Row} A = \dim \operatorname{Col} A$.

(b) $\dim \operatorname{Col} A + \dim \operatorname{Nul} A = n$.

Proof. Let s be numbers of lead variables and t the numbers of free variables of A. Then $n = s + t$. By N3.7.5, both $\operatorname{Row} A$ and $\operatorname{Col} A$ have a basis of length s, while $\operatorname{Nul} A$ has a basis of length t. □

Chapter N6

Linearity

N6.1 Basic definition

Definition 6.1. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be vector spaces and* $T : V \to W$ *a function. We say that* T *is linear function from* $\mathbf{V}$ *to* $\mathbf{W}$ *provided that*

(i) $T(u + v) = T(u) + T(v)$ *for all* $u, v \in V$, *and* *(additive)*

(ii) $T(rv) = rT(v)$ *for all* $r \in \mathbb{R}, v \in V$. *(homogeneous)*

Notation N6.1.2. *'*$T : \mathbf{V} \to \mathbf{W}$ *is linear' means that* $\mathbf{V}$ *and* $\mathbf{W}$ *are vector spaces and* T *is a linear function from* $\mathbf{V}$ *to* $\mathbf{W}$.

Theorem 6.2. *Suppose* $T : \mathbf{V} \to \mathbf{W}$ *is linear. Then*

(a) $T(\mathbf{0_V}) = \mathbf{0_W}$.

(b) $T(-v) = -T(v)$ *for all* $v \in V$.

(c) $T(r_1v_1 + r_2v_2 + \ldots + r_nv_n) = r_1T(v_1) + r_2T(v_2) + \ldots + r_nT(v_n)$ *for all list* $(v_1, \ldots, v_n)$ *in* V *and* $(r_1, \ldots, r_n)$ *in* $\mathbb{R}$.

(d) $T(u - v) = T(u) - T(v)$ *for all* $u, v \in V$.

Proof. (a): $T(\mathbf{0_V}) = T(\mathbf{0_V} + \mathbf{0_V}) = T(\mathbf{0_V}) + T(\mathbf{0_V})$ since T is linear. Thus $T(\mathbf{0_V}) = \mathbf{0_W}$ by 1.2(b).

(b): $T(v) + T(-v) = T(v + (-v)) = T(\mathbf{0}) = \mathbf{0}$ since T is linear and by (a). Thus $T(-v)$ is an additive inverse of $T(v)$ and so $T(-v) = -T(v)$ by 1.3.

(c): We prove (c) by induction. For $n = 0$, (c) says $T(\mathbf{0}) = \mathbf{0}$, which is true by (a). Suppose that (c) holds for $n = k$. Then

$$\begin{aligned}
& T(r_1v_1 + r_2v_2 + \ldots + r_{k+1}v_{k+1}) & \\
= \; & T\Big((r_1v_1 + r_2v_2 + \ldots + r_kv_k) + r_{k+1}v_{k+1}\Big) & \text{–Definition of sum of a list} \\
= \; & T(r_1v_1 + r_2v_2 + \ldots + r_kv_k) + T(r_{k+1}v_{k+1}) & \text{–Definition of linear} \\
= \; & T(r_1v_1 + r_2v_2 + \ldots + r_kv_k) + r_{k+1}T(v_{k+1}) & \text{–Definition of linear} \\
= \; & \Big(r_1T(v_1) + r_2T(v_2) + \ldots + r_kT(v_k)\Big) + r_{k+1}T(v_{k+1}) & \text{–Induction Assumption} \\
= \; & r_1T(v_1) + r_2T(v_2) + \ldots + r_kT(v_k) + r_{k+1}T(v_{k+1}) & \text{–Definition of sum of a list}
\end{aligned}$$

(d): See Homework 9 □

Lemma N6.1.4. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be vector spaces and* $T : V \to W$ *a function. Then* T *is a linear function from* $\mathbf{V}$ *to* $\mathbf{W}$ *if and only if* $T(au + bv) = a(T(u)) + b(T(v))$ *for all* $a, b \in \mathbb{R}$ *and* $u, v \in V$.

Proof. If T is linear, then by 6.3 $T(au + bv) = aT(u) + bT(v)$ for all $a, b \in \mathbb{R}$.

Suppose now that $T(au + bv) = aT(u) + bT(v)$ for all $a, b \in \mathbb{R}$ and $u, v \in V$. Choosing $a = b = 1$ we see that $T(u + v) = T(u) + T(v)$ and choosing $a = 0$ and $u = \mathbf{0}$ we see that $T(bv) = bT(v)$ for all $u, v \in V$ and $b \in \mathbb{R}$. □

Lemma N6.1.5. *Let* $\mathbf{V}$ *be a vector space and* $B = (v_1, \ldots, v_n)$ *a list in* V. *Let* $L_B : \mathbb{R}^n \to V$ *be the function defined by*

$$L_B(r_1, \ldots, r_n) = r_1v_1 + \ldots + r_nv_n$$

for all $(r_1, \ldots, r_n) \in \mathbb{R}^n$. *(In other words,* $L_B(r) = Br$ *for all* $r \in \mathbb{R}^n$.*) Then* L_B *is linear and* $L_B(e_i) = v_i$ *for all* $1 \leq i \leq n$.

Proof. Let $a, b \in \mathbb{R}$ and $x, y \in \mathbb{R}^n$. Then

$$(*) \qquad x = (r_1, \ldots, r_n) \quad \text{and } y = (s_1, \ldots, s_n).$$

for some real numbers $r_1, \ldots, r_n, s_1 \ldots, s_n$. We compute

$$\begin{aligned}
L(ax + by) &= L_B(a(r_1, \ldots, r_n) + b(s_1, \ldots, s_n)) & &\text{– (*)} \\
&= L_B(ar_1 + bs_1, \ldots, ar_n + bs_n) & &\text{– Definition of addition and} \\
& & &\quad \text{multiplication for } \mathbb{R}^n \\
&= (ar_1 + bs_1)v_1 + \ldots + (ar_n + bs_n)v_n & &\text{– definition of } L_B \\
&= a(r_1v_1 + \ldots + r_nv_n) + b(s_1v_1 + \ldots + s_nv_n) & &\text{– Axioms of a vector space} \\
&= aL_B(r_1, \ldots, r_n) + bL_B(s_1, \ldots, s_n) & &\text{– definition of } L_B \\
&= aL_B(x) + bL_B(y) & &\text{– (*)}
\end{aligned}$$

Thus L_B is linear by N6.1.4.

Also $L_B(e_i) = 0v_1 + \ldots + 0v_{i-1} + 1v_i + 0v_{i+1} + \ldots + 0v_n = \mathbf{0} + \ldots + \mathbf{0} + v_i + \mathbf{0} + \ldots + \mathbf{0} = v_i$. $\square$

Definition N6.1.6. *Let A be an $m \times n$-matrix. Then L_A is the function from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by $L_A(x) = Ax$ for all $x \in \mathbb{R}$.*

Lemma N6.1.7. *Let $n, m \in \mathbb{N}$ and A an $m \times n$-matrix. Then L_A is linear and $L_A(e_j) = a_j$ for all $1 \leq j \leq n$.*

Proof. Let $B = (a_1, a_2, \ldots, a_n)$, so B is the list of columns of A. Note that B is a list in the vector space $\mathbb{R}^m$. Let $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$. Then by the definition of L_A and L_B:

$$L_A(x) = Ax = x_1 a_1 + \ldots + x_n a_n = L_B(x).$$

Then $L_A = L_B$ (by A.2.2) and so N6.1.5 implies that L_A is linear and $L_A(e_j) = a_j$ for all $1 \leq j \leq n$. $\square$

Remark: The book uses the notation μ_A for L_A.

Theorem 6.3. *Suppose $T : \mathbf{V} \to \mathbf{W}$ and $T' : \mathbf{V} \to \mathbf{W}$ are linear.*

(a) Put $U = \{v \in V \mid T(v) = T'(v)\}$. Then U is a subspace of V.

(b) Suppose that $(v_1, \ldots, v_n)$ is a spanning list for $\mathbf{V}$ and $T(v_i) = T'(v_i)$ for all $1 \leq i \leq n$. Then $T = T'$.

Proof. (a) By 6.2(a), $T(\mathbf{0}) = \mathbf{0} = T'(\mathbf{0})$ and so $\mathbf{0} \in U$. Suppose $u, v \in U$. Then

$$\begin{aligned} T(u+v) &= T(u) + T(v) && - T \text{ is linear} \\ &= T'(u) + T'(v) && - u, v \in U \text{ and definition of } U \\ &= T'(u+v) && - T' \text{ is linear} \end{aligned}$$

Thus $T(u+v) = T'(u+v)$ and so $u + v \in U$. Now let $r \in \mathbb{R}$ and $u \in U$. Then

$$\begin{aligned} T(ru) &= rT(u) && - T \text{ is linear} \\ &= rT'(u) && - u \in U \text{ and definition of } U \\ &= T'(ru) && - T' \text{ is linear} \end{aligned}$$

Thus $T(ru) = T'(ru)$ and so $ru \in U$. We verified the three conditions of the Subspace Theorem and so U is a subspace of V.

(b) Since $T(v_i) = T'(v_i)$ we have $v_i \in U$ for all $1 \leq i \leq n$. Since $(v_1, \ldots, v_n)$ is a spanning list for V we conclude from N3.2.4 that $U = V$. So $T(v) = T'(v)$ for all $v \in V$ and thus $T = T'$ (by A.2.2). $\square$

N6.2 Composition and Inverses of functions

Definition N6.2.1. *Let I be a set and $\mathbf{V}$ a vector space.*

(a) $\mathrm{F}(I,V)$ *is the set of functions from I to V.*

(b) For $f,g\in \mathrm{F}(I,V)$ define the function $f+g$ from I to V by

$$(f+g)(i)=f(i)+g(i) \text{ for all } i\in I$$

(c) For $a\in\mathbb{R}$ and $f\in\mathrm{F}(I,V)$ define the function af from I to V by

$$(af)(i)=a(f(i)) \text{ for all } i\in I$$

(d) $\mathbf{F}(I,\mathbf{V})$ *is the triples consisting of* $\mathrm{F}(I,V)$ *and the operations in (b) and (c).*

Lemma N6.2.2. *Let I be a set and $\mathbf{V}$ a vector space.*

(a) $\mathbf{F}(I,\mathbf{V})$ *is a vector space.*

(b) The additive identity in $\mathbf{F}(I,V)$ is the zero-function 0^* defined by $0^*(v)=\mathbf{0_V}$ for all $v\in V$.

(c) The additive inverse of $f\in\mathrm{F}(I,V)$ is the function $-f$ defined by $(-f)(v)=-(f(v))$ for all $v\in V$.

Proof. See Homework 8. □

Definition N6.2.3. *Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces. Then $\mathrm{L}(\mathbf{V},\mathbf{W})$ is the set of linear functions from $\mathbf{V}$ to $\mathbf{W}$.*

Theorem N6.2.4. *Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces.*

(a) The zero function from $\mathbf{V}$ to $\mathbf{W}$ is linear.

(b) If f and g are linear function from $\mathbf{V}$ to $\mathbf{W}$, then also $f+g$ is linear.

(c) If $r\in\mathbb{R}$ and $f:\mathbf{V}\to\mathbf{W}$ is linear, then also rf is linear.

(d) $\mathrm{L}(\mathbf{V},\mathbf{W})$ *is subspace of* $\mathbf{F}(V,\mathbf{W})$.

Proof. Let $v,w\in V$ and $a,b\in\mathbb{R}$. Recall that the additive identity in $\mathbf{F}(I,V)$ is the zero function 0^* from V to W defined by $0^*(v)=\mathbf{0_W}$ for all $v\in V$. .

(a) See Homework 7 .

(b) Let $f,g:\mathbf{V}\to\mathbf{W}$ be linear. Then

$$\begin{aligned}
& (f+g)(av+bw) & \\
= \;& f(av+bw)+g(av+bw) & \text{– definition of addition of functions} \\
= \;& \Big(a\big(f(v)\big)+b\big(f(w)\big)\Big)+\Big(a\big(g(v)\big)+b\big(g(w)\big)\Big) & -f,g \text{ linear and N6.1.4} \\
= \;& a\Big(f(v)+g(v)\Big)+b\Big(f(w)+g(w)\Big) & \text{– axioms of a vector space} \\
= \;& a\Big((f+g)(v)\Big)+b\Big((f+g)(w)\Big) & \text{– definition of addition of functions, twice}
\end{aligned}$$

and so $f+g$ is linear by N6.1.4.

(c) Let $r \in \mathbb{R}$ and let $f : \mathbf{V} \to \mathbf{W}$ be linear. Then

$$\begin{aligned}
& (rf)(av+bw) & \\
= \;& r\big(f(av+bw)\big) & \text{– definition of scalar multiplication of functions} \\
= \;& r\Big(a\big(f(v)\big)+b\big(f(w)\big)\Big) & -f \text{ linear and N6.1.4} \\
= \;& a\Big(r\big(f(v)\big)\Big)+b\Big(r\big(f(w)\big)\Big) & \text{– axioms of a vector space} \\
= \;& a\Big((rf)(v)\Big)+b\Big((rf)(w)\Big) & \text{– definition of scalar multiplication of functions, twice}
\end{aligned}$$

and so by N6.1.4 rf is linear.

(d) By (a), (b) and (c) the three conditions in the Subspace Theorem hold. Thus the Subspace Theorem shows that $\mathbf{L}(\mathbf{V},\mathbf{W})$ is a subspace of $\mathbf{F}(V,\mathbf{W})$. □

Theorem 6.7. *Let $f : \mathbf{V} \to \mathbf{W}$ and $g : \mathbf{W} \to \mathbf{X}$ be linear. Then $g \circ f$ is linear.*

Proof. See Homework 9.

□

Definition 6.29. *(a) A linear function $T : \mathbf{V} \to \mathbf{W}$ is called an isomorphism if there exists a linear function $T' : \mathbf{W} \to \mathbf{V}$ with $T \circ T' = \mathrm{id}_W$ and $T' \circ T = \mathrm{id}_V$.*

(b) The vector space $\mathbf{V}$ is called isomorphic to the vector space $\mathbf{W}$ if there exists an isomorphism $T : \mathbf{V} \to \mathbf{W}$.

Theorem 6.8. *Let $f : \mathbf{V} \to \mathbf{W}$ be linear. Then the following are equivalent:*

(a) f is 1-1 and onto.

(b) f is invertible.

(c) f is invertible and $f^{-1} : \mathbf{W} \to \mathbf{V}$ is linear.

(d) f is an isomorphism.

Proof. (a) $\implies$ (b): See 6.6.

(b) $\implies$ (c): Suppose f is invertible. Let $a, b \in \mathbb{R}$ and $w, x \in W$. Put

$$(*) \qquad u = f^{-1}(w) \text{ and } v = f^{-1}(x).$$

Since f^{-1} is the inverse of f, A.5.5 gives

$$(**) \qquad w = f(u) \text{ and } x = f(v).$$

Thus

$$\begin{aligned} f^{-1}(aw+bx) &= f^{-1}\Big(a\big(f(u)\big) + b\big(f(v)\big)\Big) && -(**) \\ &= f^{-1}\Big(f(au+bv)\Big) && -f \text{ is linear and N6.1.4} \\ &= au + bv && -f^{-1} \text{ is the inverse of } f, \text{ A.5.5} \\ &= a\big(f^{-1}(w)\big) + b\big(f^{-1}(x)\big) && -(*) \end{aligned}$$

So f^{-1} is linear by N6.1.4.

(c) $\implies$ (d): Suppose f is invertible and $f^{-1} : \mathbf{W} \to \mathbf{V}$ is linear. By definition of an inverse function $f \circ f^{-1} = \mathrm{id}_W$ and $f^{-1} \circ f = \mathrm{id}_V$. By assumption f^{-1} is linear and so f is an isomorphism.

(d) $\implies$ (a): Suppose that f is an isomorphism. Then by definition there exists a linear function $f' : \mathbf{W} \to \mathbf{V}$ with $f \circ f' = \mathrm{id}_W$ and $f' \circ f = \mathrm{id}_V$. So f' is an inverse of f and by 6.6, f is 1-1 and onto. $\square$

Theorem N6.2.8. *Let* $\mathbf{V}$ *be a vector space with a basis* $B = (v_1, \ldots, v_n)$*.*

(a) Let $x \in \mathbb{R}^n$ *and* $v \in V$*. Then* $C_B(v) = x$ *if and only if* $v = L_B(x)$*.*

(b) C_B *is the inverse of* L_B*; and* L_B *is the inverse of* C_B*.*

(c) L_B *is an isomorphism from* $\mathbb{R}^n$ *to* V*; and* C_B *is an isomorphism from* V *to* $\mathbb{R}^n$*.*

(d) $C_B(v_j) = e_j$ *for all* $1 \leq j \leq n$*.*

Proof. (a) Let $v \in V$ and $x = (r_1, \ldots, r_n) \in \mathbb{R}^n$. Then

$$\begin{aligned} & C_B(v) = (r_1, \ldots, r_n) & \\ \iff \quad & [v]_B = (r_1, \ldots, r_n) & - \text{definition of } C_B \\ \iff \quad & v = r_1 v_1 + \ldots + r_n v_n & - \text{definition of } [v]_B \\ \iff \quad & v = L_B(r_1, \ldots, r_n) & - \text{definition of } L_B \end{aligned}$$

So (a) holds.

(b) follows from (a) and A.5.5.

(c) By (b) L_B is invertible and $C_B = L_B^{-1}$. By N6.1.5, L_B is linear and so by 6.8 L_B is an isomorphism and $C_B = L_B^{-1}$ is linear. Thus also C_B is an isomorphism.

(d) By N6.1.5 $L_B(e_j) = v_j$ and thus by (a) $C_B(v_j) = e_j$. □

Theorem 6.9. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be vector spaces,* $B = (v_1, \ldots, v_n)$ *a basis for* V *and* $D = (u_1, \ldots, u_n)$ *a list in* W*. Then there exists a unique linear function* $T : \mathbf{V} \to \mathbf{W}$ *with* $T(v_j) = u_j$ *for all* $1 \leq j \leq n$*, namely* $T = L_D \circ C_B$*.*

Proof. See Exercise A on Homework 10. □

N6.6 Image and Kernel

Definition 6.19. *Let* $T : \mathbf{V} \to \mathbf{W}$ *be linear. Then*

$$\ker T := \{v \in V \mid T(v) = \mathbf{0_W}\}.$$

Theorem 6.20. *Let* $T : \mathbf{V} \to \mathbf{W}$ *be linear.*

(a) $\ker T$ *is a subspace of* $\mathbf{V}$ *and* $\operatorname{Im} T$ *is a subspace of* W*.*

(b) Let $(v_1, \ldots, v_n)$ *be a spanning list for* V*. Then* $(T(v_1), \ldots, T(v_n)$ *is a spanning list for* $\operatorname{Im} T$*.*

Proof. (a) Note that $\ker T = \{v \in V \mid T(v) = 0^*(v)\}$ and so by 6.3, $\ker T$ is a subspace of V.

Since $T(\mathbf{0}) = \mathbf{0}$, $\mathbf{0} \in \operatorname{Im} T$. Let $w, x \in \operatorname{Im} T$. Then $w = T(u)$ and $x = T(v)$ for some $u, v \in V$. Thus $w + x = T(u) + T(v) = T(u + v)$ and so $w + x \in \operatorname{Im} T$. Let $r \in R$. Then $rw = rT(u) = T(ru) =$ and so $rw \in \operatorname{Im} T$. So the three conditions of the Subspace Theorem for $\operatorname{Im} T$ hold and $\operatorname{Im} T$ is a subspace of W.

(b) Let $(v_1, \ldots, v_n)$ be a spanning list for V and let $w \in W$.

$$\begin{array}{llr}
 & w \in \operatorname{Im} T & \\
\iff & w = T(v) \text{ for some } v \in V & \text{– Definition of } \operatorname{Im} T \\
\iff & w = T(r_1 v_1 + \ldots r_n v_n) \text{ for some } (r_1, \ldots, r_n) \in \mathbb{R}^n & \text{– since } (v_1, \ldots, v_n) \text{ spans} V \\
\iff & w = r_1 T(v_1) + \ldots + r_n T(v_n) \text{ for some} (r_1, \ldots, r_n) \in \mathbb{R}^n & -T \text{ is linear and 6.2(d)} \\
\iff & w \in \operatorname{span}(T(v_1), \ldots, T(v_n)) & \text{– Definition of span}
\end{array}$$

Thus $\operatorname{Im} T = \operatorname{span}(T(v_1), \ldots, T(v_n)$. □

Theorem 6.28. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear.*

(a) Let $u, v \in v$. Then $T(u) = T(v)$ if and only if $v - u \in \ker T$.

(b) T is 1-1 if and only of $\ker T = \{\mathbf{0}\}$.

Proof. (a) We have

$$\begin{aligned} & T(u) = T(v) & \\ \iff\ & T(v) - T(u) = \mathbf{0} & - 1.7(\text{p}) \\ \iff\ & T(v-u) = \mathbf{0} & -\text{T is linear and } 6.2(\text{d}) \\ \iff\ & v - u \in \ker T & - \text{Definition of } \ker T \end{aligned}$$

(b) Suppose first that T is 1-1 and let $v \in \ker T$. Then $T(v) = \mathbf{0} = T(\mathbf{0})$ and since T is 1-1, $v = \mathbf{0}$. Thus $\ker T = \{\mathbf{0}\}$.

Suppose next that $\ker T = \{\mathbf{0}\}$ and let $u, v \in V$ with $T(u) = T(v)$. Then by (a), $v - u \in \ker T$. Since $\ker T = \{\mathbf{0}\}$ this gives $v - u = \mathbf{0}$ and so by 1.7(p), $v = u$. So T is 1-1. □

Lemma N6.6.4. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear and suppose that $\mathbf{V}$ is finite dimensional. Then the following three statements are equivalent:*

(a) T is 1-1.

(b) For all linearly independent lists $(v_1, \dots, v_n)$ in V, $\big(T(v_1), \dots, T(v_n)\big)$ is linearly independent in $\mathbf{W}$.

(c) There exists a basis $(v_1, \dots, v_n)$ of $\mathbf{V}$ such that $\big(T(v_1), \dots, T(v_n)\big)$ is linearly independent in $\mathbf{W}$.

Proof. (a) $\Longrightarrow$ (b): (a) Suppose T is 1-1 and let $(r_1, \dots, r_n) \in \mathbb{R}^n$ with

$$r_1 T(v_1) + \ldots + r_n T(v_n) = \mathbf{0}.$$

Using 6.2 this gives:

$$T(r_1 v_1 + \ldots + r_n v_n) = T(\mathbf{0}),$$

and since T is 1-1, we get

$$r_1 v_1 + \ldots + r_n v_n = \mathbf{0}.$$

Since $(v_1, \dots, v_n)$ is linearly independent this implies $r_1 = r_2 = \ldots = r_n = 0$. So $\big(T(v_1), \dots, T(v_n)\big)$ linearly independent.

(b) $\Longrightarrow$ (c): By N3.4.4 V has a basis $B = (v_1, \dots, v_n)$. Then B is linearly independent and so by (b) $(T(v_1), \dots, T(v_n))$ is linearly independent.

(c) $\implies$ (a): Suppose $B = (v_1, \ldots, v_n)$ is basis for V such that $(T(v_1), \ldots, T(v_n))$ is linearly independent in $\mathbf{W}$. Let $v \in \ker T$. Since B spans V, $v = r_1v_1 + \ldots + r_nv_n$ for some $(r_1, \ldots, r_n) \in \mathbb{R}^n$. Using 6.2 we get

$$r_1T(v_1) + \ldots + r_nT(v_n) = T(r_1v_1 + \ldots + r_nv_n) = T(v) = \mathbf{0}$$

and since $(T(v_1), \ldots, T(v_n))$ is linearly independent, $r_1 = 0, r_2 = 0, \ldots, r_n = 0$. Thus $v = \mathbf{0}$ and so $\ker T = 0$. Thus by 6.28 T is 1-1. □

Lemma N6.6.5. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear and suppose that $\mathbf{V}$ is finite dimensional. Then the following three statements are equivalent:*

(a) T is onto.

(b) Im $T = \mathbf{W}$.

(c) For all spanning list $(v_1, \ldots, v_n)$ of V, $\big(T(v_1), \ldots, T(v_n)\big)$ spans W.

(d) There exists a spanning list $(v_1, \ldots, v_n)$ of V such that $\big(T(v_1), \ldots, T(v_n)\big)$ spans W.

Proof. (a) $\iff$ (b) : By definition T is onto if and only if Im $T = W$. So (a) and (b) are equivalent.

(b) $\implies$ (c): Suppose Im $T = \mathbf{W}$. Let $(v_1, \ldots, v_n)$ be spanning list for V. By 6.20(b) Im $T = \operatorname{span}\big(T(v_1), \ldots, T(v_n)\big)$ and since Im $T = \mathbf{W}$ we conclude that $\big(T(v_1), \ldots, T(v_n)\big)$ spans $\mathbf{W}$.

(c) $\implies$ (d): Suppose (c) holds. Since $\mathbf{V}$ is finite dimensional, $\mathbf{V}$ has a spanning list $(v_1, \ldots, v_n)$. Since (c) holds we conclude that $\big(T(v_1), \ldots, T(v_n)\big)$ spans W. Thus (d) holds.

(d) $\implies$ (b): Suppose there exists a spanning list $(v_1, \ldots, v_n)$ of V such that

$$(*) \qquad \big(T(v_1), \ldots, T(v_n)\big) \text{ spans } W$$

By 6.20(b) Im $T = \operatorname{span}\big(T(v_1), \ldots, T(v_n)\big)$ and so by (*) Im $T = W$. □

Corollary N6.6.6. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear and suppose $\mathbf{V}$ is finite dimensional. Then the following are equivalent.*

(a) T is invertible.

(b) For all basis $(v_1, \ldots, v_n)$ of $\mathbf{V}$, $\big(T(v_1), \ldots, T(v_n)\big)$ is a basis for V.

(c) There exists a basis $(v_1, \ldots, v_n)$ of $\mathbf{V}$ such that $\big(T(v_1), \ldots, T(v_n)\big)$ is a basis for V.

Proof. By 6.6, T is invertible if and only if T is 1-1 and onto. By definition $\big(T(v_1), \ldots, T(v_n)\big)$ is basis for W if and only if $\big(T(v_1), \ldots, T(v_n)\big)$ is linearly independent and spans W. Thus the corollary follows from N6.6.4 and N6.6.5 □

Corollary N6.6.7. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear and suppose $\mathbf{V}$ and $\mathbf{W}$ are finite dimensional and $\dim \mathbf{V} = \dim \mathbf{W}$. Then the following are equivalent:*

(a) T is invertible.

(b) T is 1-1.

(c) T is onto.

Proof. Let $n = \dim \mathbf{V} = \dim \mathbf{W}$ and let $(v_1, \dots, v_n)$ a basis of $\mathbf{V}$. Then

$$\begin{array}{clr} & T \text{ is 1-1} & \\ \iff & \big(T(v_1), \dots, T(v_n)\big) \text{ is linearly independent} & -\text{N6.6.4} \\ \iff & \big(T(v_1), \dots, T(v_n)\big) \text{ spans } W & -N3.5.5 \\ \iff & T \text{ is onto} & -\text{N6.6.5} \end{array}$$

In particular, T is 1-1 if and only if T is 1-1 and onto, and so by 6.6 if and only if T is invertible. □

Corollary N6.6.8. *Let $\mathbf{V}$ and $\mathbf{W}$ be finite dimensional vector spaces of equal dimension. Let $T : \mathbf{V} \to \mathbf{W}$ and $S : \mathbf{W} \to \mathbf{V}$ be linear. Then the following four statements are equivalent.*

(a) $S \circ T = \mathrm{id}_V$.

(b) S is an inverse of T.

(c) T is an inverse of S

(d) $T \circ S = \mathrm{id}_W$.

Proof. (a) $\implies$ (b): Suppose that $S \circ T = \mathrm{id}_V$. Then by Homework Problem 6.2(9b) T is 1-1. Since $\mathbf{V}$ and $\mathbf{W}$ have equal dimension N6.6.7 shows that T is invertible. Since $S \circ T = \mathrm{id}_V$ we conclude that $S = \mathrm{id}_V \circ T^{-1} = T^{-1}$, see A.5.3 .

(b) $\implies$ (c): If S is an inverse of T, then T is also an inverse of S, see A.5.5

(c) $\implies$ (d): If T is an inverse of S, then $T \circ S = \mathrm{id}_W$ by definition of an inverse.

(d) $\implies$ (a): Suppose $T \circ S = \mathrm{id}_W$. The result that (a) implies (d), applied with the roles of T and S interchanged, shows that T is an inverse of S. Thus $S \circ T = \mathrm{id}_V$, by definition of an inverse.

□

Definition 5.3. *Let n be a non-negative integer. I_n is the $n \times n$-matrix $[e_1, \ldots, e_n]$, so column j of I_n is e_j and*

$$I_n = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 & 0 \\ 0 & 1 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & 1 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & 1 & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & 1 & 0 \\ 0 & 0 & 0 & \ldots & 0 & 0 & 1 \end{bmatrix}$$

We will often just write I for I_n.

Definition 5.1. *Let A be an $m \times n$-matrix and B an $n \times p$ matrix. Then AB is the $m \times p$-matrix whose jth column is equal to Ab_j. So*

$$A = [Ab_1, Ab_2, \ldots, Ab_p].$$

We denote column j of AB by $(ab)_j$ and entry (i,j) of AB by $(ab)_{ij}$. So $(ab)_j = Ab_j$ and $(ab)_{ij}$ is entry i of Ab_j. Hence

$$(ab)_{ij} = a^i b_j = a_{i1}b_{1j} + a_{i2}b_{j2} + \ldots + a_{in}b_{nj},$$

and

$$AB = \left[a_{i1}b_{1j} + a_{i2}b_{j2} + \ldots + a_{in}b_{nj}\right]_{i=1,j=1}^{m,p}.$$

Definition 5.6. *Let n and m be positive integers and A an $m \times n$ matrix.*

(a) An inverse of A is an $n \times m$ matrix B with

$$AB = I_n \text{ and } BA = I_m$$

(b) A is called invertible if A has an inverse.

Theorem 6.22. *Let n and m be positive integer and A an $m \times n$-matrix. Then*

(a) $\mathrm{Col}A = \mathrm{Im}\ L_A$.

(b) $\mathrm{Nul}A = \ker L_A$.

Proof. (a) Since $e_1, \ldots, e_n$ spans $\mathbb{R}^n$, 6.20(b) implies that $(L_A(e_1), \ldots, L_A(e_n))$ spans $\mathrm{Im}\ L_A$. Since $L_A(e_j) = Ae_j = a_j$, we get $\mathrm{Im}\ L_A = \mathrm{span}(a_1, \ldots, a_n) = \mathrm{Col}A$ and (a) holds.

(b) Let $x \in \mathbb{R}^n$. Then $L_A(x) = Ax$ and so $x \in \ker L_A$ if and only if $Ax = \mathbf{0}$. $\square$

Lemma N6.6.13. *Let n and m be positive integer and A an $m \times n$-matrix. Then the following are equivalent:*

(a) The list of columns of A is linearly independent.

(b) $\mathrm{Nul}A = \{\mathbf{0}\}$.

(c) L_A is 1-1.

Proof. We have

$$\begin{array}{lcr}
 & \mathrm{Nul}A = \{\mathbf{0}\} & \\
\iff & \mathrm{Nul}L_A = \{\mathbf{0}\} & -\ 6.22 \\
\iff & L_A \text{ is 1-1} & -\ 6.28 \\
\iff & (L_A(e_1), \ldots, L_A(e_n)) \text{ is linearly independent} & -\ \text{N6.6.4} \\
\iff & (a_1, \ldots, a_n) \text{ is linearly independent} & -\ \text{N6.1.7}
\end{array}$$

□

Lemma N6.6.14. *Let n and m be positive integer and A an $m \times n$-matrix. Then the following are equivalent:*

(a) The list of columns of A spans $\mathbb{R}^m$

(b) $\mathrm{Col}A = \mathbb{R}^m$.

(c) $\mathrm{Im}\ L_A = \mathbb{R}^m$.

(d) L_A is onto

Proof. Since $\mathrm{Col}A$ is the span of the columns of A, (a) and (b) are equivalent. Since $\mathrm{Col}A = \mathrm{Im}\ L_A$, (b) and (d) are equivalent. Finally, by definition of onto, L_A is onto if and only if $\mathrm{Im}\ L_A = \mathbb{R}^m$ and so (d) and (c) are equivalent. □

N6.3 Matrix of a Linear Function

Definition 6.12. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear. Suppose $B = (v_1, \ldots, v_n)$ is a basis for $\mathbf{V}$ and $D = (w_1, \ldots, w_m)$ is a basis for $\mathbf{W}$. Let A be the $m \times n$ matrix with $a_j = [T(v_j)]_D$ for all $1 \leq j \leq n$. So*

$$A = \Big[[T(v_1)]_D, [T(v_2)]_D, \ldots, [T(v_n)]_D \Big].$$

Then A is called the matrix of T with respect to B and D.

Lemma N6.3.2. *Let $\mathbf{V}$ and $\mathbf{W}$ be vector spaces with bases $B = (v_1, \ldots, v_n)$ and $D = (w_1, \ldots, w_m)$ respectively.*

(a) $(L_A \circ C_B)(v) = A[v]_B$ *for all* $v \in V$.

(b) Let $T : V \to W$ *be a function then* $(C_D \circ T)(v) = [T(v)]_D$ *for all* $v \in V$.

Proof. Let $v \in V$. (a)

$$\begin{aligned} & (L_A \circ C_B)(v) & \\ = \; & L_A(C_B(v)) & \text{–definition of composition} \\ = \; & L_A\big([v]_B\big) & \text{–definition of } C_B \\ = \; & A[v]_B & \text{–definition of } L_A \end{aligned}$$

(b)

$$\begin{aligned} & (C_D \circ T)(v) & \\ = \; & C_D(T(v)) & \text{–definition of composition} \\ = \; & [T(v)]_D & \text{–definition of } C_D \end{aligned}$$

□

Theorem 6.11. *Let* **V** *and* **W** *be vector spaces with bases* $B = (v_1, \ldots, v_n)$ *and* $D = (w_1, \ldots, w_m)$ *respectively. Put* $n = \dim V$ *and* $m = \dim W$ *and let* $A \in \mathbb{M}(m, n)$. *Let* $T : V \to W$ *be a function. Then the following are equivalent*

(a) T *is linear and* A *is the matrix of* T *with respect to* B *and* D.

(b) $C_D \circ T = L_A \circ C_B$.

(c) $C_D \circ (T \circ L_B) = L_A$.

(d) $T \circ L_B = L_D \circ L_A$.

(e) $T = L_D \circ (L_A \circ C_B)$.

(f) $[T(v)]_D = A[v]_B$ *for all* $v \in V$.

(g) $T(r_1v_1 + \ldots + r_nv_n) = (a_{11}r_1 + \ldots + a_{1n}r_n)w_1 + \ldots + (a_{m1}r_1 + \ldots + a_{mn}r_n)w_m$ *for all* $(r_1, \ldots, r_n) \in \mathbb{R}^n$.

The functions appearing in the theorem can be visualized in the following diagram

$$\begin{tikzcd} V \arrow[r, "T"] \arrow[d, shift left, "C_B"] & W \arrow[d, shift right, "C_D"'] \\ \mathbb{R}^n \arrow[u, shift left, "L_B"] \arrow[r, "L_A"'] & \mathbb{R}^n \arrow[u, shift right, "L_D"'] \end{tikzcd}$$

Proof. Let $B = (v_1, \ldots, v_n)$ and $1 \leq j \leq n$.

$$(*) \qquad \begin{array}{rll} & (L_A \circ C_B)(v_j) & \\ = & L_A(C_B(v_j)) & -\text{ definition of composition} \\ = & L_A(e_j) & -\text{ N6.2.8} \\ = & a_j & -\text{N6.1.7} \end{array}$$

(a) $\Longrightarrow$ (b): Suppose that T is linear and A is the matrix of T (with respect to B and D.) Then

$$\begin{array}{rll} & (C_D \circ T)(v_j) & \\ = & [T(v_j)]_D & -\text{N6.3.2(a)} \\ = & a_j & -\text{ definition of } A \end{array}$$

and so by (*) $(C_D \circ T)(v_j) = a_j = (L_A \circ C_B)(v_j)$.

Since T, L_A, C_D and C_B are linear, also $L_A \circ C_B$ and $C_D \circ T$ are linear by 6.7. Since B spans V, we conclude from 6.3 that $C_D \circ T = L_A \circ C_B$. So (b) holds.

We will now show that (b), (c), (d) and (e) are equivalent.

$$\begin{array}{rll} & C_D \circ T = L_A \circ C_B & \\ \Longleftrightarrow & T = L_D \circ (L_A \circ C_B) & -L_D = C_D^{-1} \text{ and A.5.3(b)} \\ \Longleftrightarrow & T = (L_D \circ L_A) \circ C_B & -\text{ composition is associative} \\ \Longleftrightarrow & T \circ L_B = L_D \circ L_A & -L_B = C_B^{-1} \text{ and A.5.3(f)} \\ \Longleftrightarrow & C_D \circ (T \circ L_B) = L_A & -L_D = C_D^{-1} \text{ and A.5.3(b)} \end{array}$$

(e) $\Longrightarrow$ (a): Suppose that (e) holds. Then $T = L_D \circ L_A \circ C_B$. Since L_D, L_A and C_B are linear and composition of linear functions are linear (see 6.7) also $T = L_D \circ L_A \circ C_B$ is linear. We compute

$$\begin{array}{rll} & T(v_j) & \\ = & (L_D \circ L_A \circ C_B)(v_j) & -\text{ since (e) holds} \\ = & L_D\Big((L_A \circ C_B)(v_j)\Big) & \text{definition of composition} \\ = & L_D(a_j) & -\ (*) \end{array}$$

We proved $T(v_j) = L_D(a_j)$ and so by N6.2.8(a), $C_D(T(v_j)) = a_j$. The definition of C_B gives $[T(v_j)]_D = a_j$. Hence A is the matrix of T with respect to B and D.

(b) $\iff$ (f) : We have

$$\begin{array}{lll} & C_D \circ T = L_A \circ C_B & \\ \iff & (C_D \circ T)(v) = (L_A \circ C_B)(v) \text{ for all } v \in V & - \text{A.2.2} \\ \iff & [T(v)]_D = A[v]_B \text{ for all } v \in V & - \text{N6.3.2(a) and (b)} \end{array}$$

Thus (b) and (f) are equivalent.
(d) $\iff$ (g) :
We compute

$$(**) \qquad \begin{array}{lll} & (T \circ L_B)(r_1, \ldots, r_n) & \\ = & T\big(L_B(r_1, \ldots, r_n)\big) & - \text{Definition of composition} \\ = & T(r_1 v_1 + \ldots + r_n v_n) & - \text{Definition of } L_B \end{array}$$

and

$(***)$

$$\begin{array}{lll} & (L_D \circ L_A)(r_1, \ldots r_n) & \\ = & L_D\big(L_A(r_1, \ldots, r_n)\big) & - \text{Definition of composition} \\ = & L_D\big(A(r_1, \ldots, r_n)\big) & - \text{Definition of } L_A \\ = & L_D(a_{11}r_1 + \ldots a_{1n}r_n, \ldots, a_{m1}r_1 + \ldots a_{mn}r_n) & - \text{Definition of } Ax \\ = & (a_{11}r_1 + \ldots + a_{1n}r_n)w_1 + \ldots + (a_{m1}r_1 + \ldots + a_{mn}r_n)w_m & - \text{Definition of } L_D \end{array}$$

Thus

$$\begin{array}{lll} & T \circ L_B = L_D \circ L_A & \\ \iff & (T \circ L_B)(r_1, \ldots, r_n) = (L_A \circ L_B)(r_1, \ldots, r_n) \text{ for all } (r_1, \ldots, r_n) \in \mathbb{R}^n & - \text{A.2.2} \\ \iff & T(r_1 v_1 + \ldots + r_n v_n) = (a_{11}r_1 + \ldots + a_{1n}r_n)w_1 + \ldots + (a_{m1}r_1 + \ldots + a_{mn}r_n)w_m & - (**) \text{ and } (***) \\ & \text{for all } (r_1, \ldots, r_n) \in \mathbb{R}^n & \end{array}$$

Thus (d) and (g) are equivalent.

□

Theorem 6.10. *Let n and m be positive integers, $A \in \mathbb{M}(m,n)$ and $T : \mathbb{R}^n \to \mathbb{R}^m$ a function. Then the following two statements are equivalent*

(a) T is linear and A is the matrix of T with respect the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

(b) $T = L_A$.

(c) $T(x) = Ax$ *for all* $x \in \mathbb{R}^n$*.*

Proof. Let B and D be the standard basis for $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively. By 6.11(a), (b) and (f) the following three statements are equivalent:

(a') T is linear and A is the matrix of T with respect the standard bases of $\mathbb{R}^n$ and $\mathbb{R}^m$.

(b') $C_D \circ T = L_A \circ C_B$.

(c') $[T(x)]_D = A[x]_B$ for all $x \in \mathbb{R}^n$.

Observe that (a') is (a).

Also since B and D are the standard bases, $[x]_B = x$ and $[y]_D = y$ for all $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. So (c') is equivalent to (c).

Moreover, $C_B = \mathrm{id}_{\mathbb{R}^n}$ and $C_D = \mathrm{id}_{\mathbb{R}^m}$ and so $C_D \circ T = T$ and $L_A \circ C_B = L_A$. Thus (b') is equivalent to (b) and the theorem is proved. □

N6.4 The matrices of composition and inverses

Theorem N6.4.1. *(a) Let* I *be a set,* $\mathbf{V}$ *a vector space,* $f, g : I \to \mathbf{V}$ *functions and* $i \in I$*. Then*

$$(f \pm g)(i) = f(i) \pm g)(i)$$

(b) Let I *and* J *be sets and* $\mathbf{V}$ *a vector space. Let* $f : I \to J$ *and* $g, h : J \to V$ *be functions and let* $r \in \mathbb{R}$*. Then*

$$(g + h) \circ f = g \circ f \pm h \circ f \quad \textit{and} \quad g \circ (rf) = r(g \circ f)$$

(c) Let I *be a set and* $\mathbf{V}$ *and* $\mathbf{W}$ *vector spaces. Let* $f, g : I \to \mathbf{V}$ *be functions and* $h : \mathbf{V} \to \mathbf{W}$ *a linear function and let* $r \in \mathbb{R}$*. Then*

$$h \circ (f + g) = h \circ f \pm h \circ g \quad \textit{and} \quad h \circ (rf) = r(h \circ f).$$

Proof. Let $i \in I$.

(a) $(f + g)(i) = f(i) + g(i)$ holds by the definition of addition of functions.

$$\begin{aligned}
& (f-g)(i) \\
= \;& (f+(-g))(i) && - \text{ definition of subtraction} \\
= \;& f(i)+(-g)(i) && - \text{ definition of addition of functions} \\
= \;& f(i)+(-(g(i))) && - \text{ N6.2.2(c)} \\
= \;& f(i)-g(i) && - \text{ definition of subtraction}
\end{aligned}$$

(b) We have

$$
\begin{aligned}
& \Big((g \pm h) \circ f\Big)(i) \\
= \;& \Big(g \pm h\Big)\big(f(i)\big) && \text{– definition of composition} \\
= \;& g\big(f(i)\big) \pm h\big(f(i)\big) && \text{–(a)} \\
= \;& (g \circ f)(i) \pm (h \circ f)(i) && \text{– definition of composition,twice} \\
= \;& (g \circ f \pm h \circ f)(i) && \text{–(a)}
\end{aligned}
$$

and so the first statement in (b) holds by A.2.2.

Also

$$
\begin{aligned}
& \Big((rg) \circ f\Big)(i) \\
= \;& \big(rg\big)\big(f(i)\big) && \text{– definition of composition} \\
= \;& r\Big(g\big(f(i)\big)\Big) && \text{– definition of scalar multiplication of functions} \\
= \;& r\Big((g \circ f)(i)\Big) && \text{– definition of composition} \\
= \;& \Big(r(g \circ f)\Big)(i) && \text{– definition of scalar multiplication of functions}
\end{aligned}
$$

and so the second statement in (b) holds by A.2.2.

(c) We have

$$
\begin{aligned}
& \Big(h \circ (f \pm g)\Big)(i) \\
= \;& h\Big((f \pm g)(i)\Big) && \text{– definition of composition} \\
= \;& h\Big(f(i) \pm g(i)\Big) && \text{– (a)} \\
= \;& h\big(f(i)\big) \pm h\big(g(i)\big) && \text{– since } h \text{ is linear} \\
= \;& (h \circ f)(i) \pm (h \circ g)(i) && \text{– definition of composition,twice} \\
= \;& (h \circ f \pm h \circ g)(i) && \text{– (a)}
\end{aligned}
$$

and so the first statement in (c) holds by A.2.2. Also

$$\begin{aligned}
&\Big((h \circ (rf)\Big)(i) \\
= \;& h\Big((rf)(i)\Big) && \text{– definition of composition} \\
= \;& h\Big(r\big(f(i)\big)\Big) && \text{– definition of scalar multiplication of functions} \\
= \;& r\Big(h\big(f(i)\big)\Big) && \text{– since } h \text{ is linear} \\
= \;& r\big((h \circ f)(i)\big) && \text{– definition of composition} \\
= \;& \Big(r(h \circ f)\Big)(i) && \text{– definition of scalar multiplication of functions}
\end{aligned}$$

and so the second statement in (c) holds by A.2.2. □

Lemma N6.4.2. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be vector spaces with bases* B *and* D *respectively. Let* $T : \mathbf{V} \to \mathbf{W}$ *be linear and let* A *be the matrix of* T *with respect to* B *and* D

(a) Let $T' : \mathbf{V} \to \mathbf{W}$ *be linear and let* A' *be the matrix of* T' *with respect to* B *and* D*. Then* $A \pm A'$ *is the matrix of* $T \pm T'$ *with respect to* B *and* D*.*

(b) Let $r \in \mathbb{R}$*. Then* rA *is the matrix of* rT *with respect to* B *and* D*.*

(c) Let $T' : \mathbf{W} \to \mathbf{U}$ *be linear,* E *a basis for* $\mathbf{U}$ *and* A' *the matrix for* T' *with respect to* D *and* E*. Then* $A'A$ *is the matrix for* $T' \circ T$ *with respect to* B *and* E

(d) Let $T' : \mathbf{V} \to \mathbf{V}$ *be linear and* A' *the matrix for* T' *with respect to* B *and* B*. Then* $T' = \mathrm{id}_V$ *if and only if* $A' = I$*.*

(e) Let $T' : \mathbf{W} \to \mathbf{V}$ *be linear and* A' *the matrix for* T' *with respect to* D *and* B*. Then* $T' \circ T = \mathrm{id}_V$ *if and only if* $A'A = I$*. In particular,* T' *is an inverse of* T *if and only if* A' *is an inverse of* A*.*

Proof. Let $B = (v_1, v_2, \ldots, v_n)$.

(a) Since C_D is linear, N6.4.1(c) gives

$$C_D \circ (T \pm T') = C_D \circ T \pm C_D \circ T'.$$

So Column j of the matrix of $T \pm T'$ is

$$\begin{aligned}
\Big[(T \pm T')(v_j)\Big]_D &= \Big(C_D \circ (T \pm T')\Big)(v_j) &&= (C_D \circ T)(v_j) \pm (C_D \circ T')(v_j) \\
&= [T(v_j)] \pm [T'(v_j)]_D &&= a_j \pm a'_j
\end{aligned}$$

Since $a_j \pm a'_j$ also is Column j of $A \pm A'$, (a) holds.

(b) Since C_D is linear, N6.4.1(c) gives $C_D \circ (rT) = r(C_D \circ T)$. So Column j of the matrix of rT is

$$\Big[(rT)(v_j)\Big]_D = \Big(C_D \circ (rT)\Big)(v_j) = r\Big((C_D \circ T)(v_j)\Big) = r[T(v_j]_D = ra_j.$$

Since ra_j is also Column j if rA, (b) holds.

(c) Note that $[T(v_j)]_D = a_j$. So by 6.11 $\Big[T'\big(T(v_j)\big)\Big]_E = A'a_j$ and so $\Big[(T' \circ T)(v_j)\Big]_E = A'a_j$. Thus Column j of the matrix of $T' \circ T$ is $A'a_j$, which also is Column j of $A'A$. So (c) is proved.

(d) Suppose that $T' = \mathrm{id}_V$. Then

$$a'_j = [T(v_j)]_B = [\mathrm{id}_V(v_j)]_B = [v_j]_B = e_j$$

and so $A = I_n$. Suppose that $A' = I_n$. We just proved that also the matrix for id_V is I_n. By 6.11 a linear function is uniquely determined by its matrix and so $T' = \mathrm{id}_V$.

(e) By (c), $A'A$ is the matrix of $T' \circ T$ with respect to B and B and so by (d), $T' \circ T = \mathrm{id}_V$ if and only if $A'A = I$. By symmetry, $T \circ T' = \mathrm{id}_W$ if and only if $AA' = I$. Thus (e) holds. $\square$

Chapter N5

Matrices

N5.1 Matrix Algebra

Theorem 5.4. *Let $n, m, p, q \in \mathbb{N}$, $A, A' \in M(m,n)$, B, B' in $M(n,p)$, $C \in M(p,q)$ and $r \in \mathbb{R}$. Then*

(a) $A(BC) = (AB)C)$

(b) $(A \pm A')B = AB \pm A'B.$

(c) $A(B \pm B') = AB \pm AB'.$

(d) $(rA)B = r(AB) = A(rB).$

(e) $AI_n = A = I_m A.$

Proof. By 6.10 an $m \times n$ matrix is essentially the same as a linear function from $\mathbb{R}^n$ to $\mathbb{R}^m$. Together with N6.4.2 the statements in this theorem follow easily from the corresponding results for linear function. As an example we will prove (a) and (d); and leave the proofs of the remaining statements to the reader.

By 6.10 A, B and C are the matrices of L_A, L_B and L_C,respectively, with respect to the standard bases.

(a) By N6.4.2 the matrices of $L_A \circ (L_B \circ L_C)$ and $(L_A \circ L_B) \circ L_C$ are $A(BC)$ and $(AB)C$, respectively. By A.4.3 $L_A \circ (L_B \circ L_C) = (L_A \circ L_B) \circ L_C$ and so $(AB)C = A(BC)$.

(d) By N6.4.2 the matrices of $(rL_A) \circ L_B)$, $r(L_A \circ L_B)$ and $L_A \circ (rL_B)$ are $(rA)B$, $r(AB)$ and $A(rB)$ respectively.

By 6.7 $(rL_A) \circ L_B) = r(L_A \circ L_B) = L_A \circ (rL_B)$ and so $(rA)B = r(AB) = A(rB)$. □

N5.2 Inverses

Lemma N5.2.1. *Let A be ab $n \times n$ matrix. Then the following statements are equivalent.*

(a) A is invertible.

(b) L_A *is invertible.*

(c) L_A *is 1-1,*

(d) L_A *is onto.*

(e) $\mathrm{Nul}A = \{\mathbf{0}\}$.

(f) The list of columns of A *is linearly independent.*

(g) $\mathrm{Col}A = \mathbb{R}^n$.

(h) The list of columns of A *spans* $\mathbb{R}^n$.

(i) The list of columns of A *is basis of* $\mathbb{R}^n$.

(j) $\dim \mathrm{Col}A = n$.

Proof. By N6.4.2(e) (a) and (b) are equivalent.

By N6.6.7 (b), (c) and (d) are equivalent.

By N6.6.13 (c), (e) and (f) are equivalent.

By N6.6.14 (d), (g) and (h) are equivalent.

Since (f) and (h) are equivalent, they are also equivalent to (i).

By a homework problem, $\mathrm{Col}A = \mathbb{R}^n$ if and only if $\dim \mathrm{Col}A = n$. So (g) and (j) are equivalent. □

Remark: Since $\dim \mathrm{Col}A = \dim \mathrm{Row}A$ the preceding theorems stays true for rows in place of columns.

Theorem 5.13. *Let* A *and* B *be* $n \times n$*-matrices. Then* $AB = I_n$ *if and only if* $BA = I$ *and if and only if* B *is an inverse of* A.

Proof. We have

$$\begin{aligned}
& AB = I_n \\
\iff\quad & L_{AB} = \mathrm{id}_{\mathbb{R}^n} && - \text{N6.4.2(d)} \\
\iff\quad & L_A \circ L_B = \mathrm{id}_{\mathbb{R}^n} && - \text{N6.4.2(c)} \\
\iff\quad & L_B \circ L_A = \mathrm{id}_{\mathbb{R}^n} && - \text{N6.6.8} \\
\iff\quad & L_{BA} = \mathrm{id}_{\mathbb{R}^n} && - \text{N6.4.2(c)} \\
\iff\quad & BA = I_n && - \text{N6.4.2(d)}
\end{aligned}$$

□

Lemma N5.2.3. *Let* B *be a* $m \times n$ *matrix in reduced row-echelon form. Then the following are equivalent.*

(a) Each column of B contains a leading 1.

(b) B has exactly n non-zero rows.

(c) B has at least n-non zero rows.

(d) $b_j = e_j$ for all $1 \leq j \leq n$, where $(e_1, \ldots, e_m)$ is the standard basis for $\mathbb{R}^m$.

Proof. Let s be the number of lead variables of A and x_{l_i} the i'th lead variable. Then $l_1 < l_2 < \ldots < l_s$ and by N3.7.5(a) $b_{l_i} = e_i$. Thus B has exactly s columns containing leading 1's. So $s \leq n$ and (a), (b) and (c) are equivalent.

If $B = I_n$, then all rows of B are non-zero and so (d) implies (b).

Suppose now that each column of B contains a leading 1. Then $t = n$ and so $(l_1, \ldots, l_s) = (1, \ldots, n)$. Thus $l_i = i$ and so $b_i = b_{l_i} = e_i$. hence (a) implies (d). □

Lemma N5.2.4. *Let A be $n \times n$ matrix.*

(a) A is invertible if and only if the reduced echelon form of A is I_n.

(b) If A is invertible and P is an $m \times m$ matrix, the reduced row-echelon form of $[A, P]$ is $[I_n, A^{-1}P]$.

(c) If A is invertible, the reduced row-echelon form of $[A, I_n]$ is $[I_n, A^{-1}]$.

Proof. (a) Let B be the reduced echelon form of A and let t be the number of lead variables of B. Let N5.2.1 A is invertible if and only if $\dim \mathrm{Col} A = n$. By N3.7.5 $\dim \mathrm{Col} A = t$ is the number of lead variables. So A is invertible if and only if $t = n$. Since F is an $n \times n$ matrix, this holds if and only row and columns contains a leading one. By N5.2.3 this holds if and only if $B = I_n$.

(b) Let $D = [A, P$ and F the reduced echelon form of D. Then $F = [B, H]$ for some $n \times n$ matrices B and H. Note that B is the reduced echelon form of A and so by (a) $B = I_n$. Thus $b_k = e_k$ for all $1 \leq k \leq n$. Let $1 \leq j \leq m$, then

$$f_{n+j} = h_j = h_{1j}e_1 + \ldots h_{nj}e_n = h_{1j}b_1 + \ldots h_{nj}b_n = h_{1j}f_1 + \ldots h_{nj}f_n$$

Thus by N3.7.4(a)

$$p_j = d_{n+j} = h_{1j}d_1 + \ldots h_{nj}d_n = h_{1j}a_1 + \ldots h_{nj}a_n = Ah_j$$

Hence $P = AH$ and so by 5.13 so $A^{-1}P = A^{-1}(AP) = (A^{-1}A)P = IP = P$. . Thus $F = [B, H] = [I_n, A^{-1}P]$.

(c) Since $A^{-1}I = A^{-1}$, this follows from (b) applied with $P = I_n$. □

Chapter N6

Linearity (Cont.)

N6.5 Change of basis

Definition 6.16. *Let V be a vector space with basis B and B'. Then the change-of-basis matrix from B' to B is the matrix of* id_V *with respect to B' and B.*

Theorem 6.15. *Let V be a vector space with basis B' and B and P the change-of-basis matrix from B' to B. Then*

(a) $[v]_B = P[v]_{B'}$ for all $v \in V$. That is the diagram

$$\begin{tikzcd} & V \arrow[dl, "C_{B'}"'] \arrow[dr, "C_B"] & \\ \mathbb{R}^n \arrow[rr, "L_P"] & & \mathbb{R}^n \end{tikzcd}$$

commutes.

(b) If $B' = (v'_1, \ldots, v'_n)$, then $p_j = [v'_j]_B$. So

$$P = \Big[[v'_1]_B, [v'_2]_B, \ldots, [v'_n]_B \Big]$$

Proof. (a) Since P is the matrix of id_V, 6.11 gives $[v]_B = [\mathrm{id}_V(v)]_B = P[v]_{B'}$.

(b) By definition of P, column j of P is $[\mathrm{id}_V(v'_j)]_B$, which is equal to $[v'_j]_B$. $\square$

Theorem 6.17. *Let V be a vector space with basis B' and B, and let P be the change-of-basis matrix from B' to B. Then P is invertible and P^{-1} is the change of basis matrix from B to B'.*

Proof. Note that id_V is invertible with inverse id_V. By definition, P is the matrix of id_V with respect to B' and B. So by N6.4.2(e), P is invertible and P^{-1} is the matrix of id_V with respect to B and B'. Hence P^{-1} is the change-of-basis matrix from B to B'. $\square$

Lemma N6.5.4. *Let V be vector space with basis $B = (v_1, \ldots, v_n)$. Let P an invertible $n \times n$ matrix and put $v'_j = L_B(p_j)$ and $B' = (v'_1, \ldots, v'_n)$. Then B' is a basis for V and P is the change-of-basis matrix from B' to B.*

Proof. Since P is invertible, $(p_1, \ldots, p_n)$ is a basis for $\mathbb{R}^n$ by N5.2.1. By N6.2.8 L_B is invertible (with inverse C_B) and so by N6.6.6 B' is basis for V. Moreover, $[v'_j]_B = C_B\Big(L_B(p_j)\Big) = p_j$ and so by 6.15(b), P is the change-of-basis matrix from B' to B. $\square$

Lemma 6.18. *Let $T : \mathbf{V} \to \mathbf{W}$ be linear. Let B and B' be basis for $\mathbf{V}$ and let D and D' be bases for $\mathbf{W}$. Suppose:*

(i) A is the matrix of T with respect to B and D.

(ii) A' is the matrix of T with respect to B' and D'.

(iii) P is the change-of-basis matrix from B' to B.

(iv) Q is the change-of-basis matrix from D' to D.

Then

$$A' = Q^{-1}AP$$

Proof. Note by 6.17 Q^{-1} is the matrix of id_W with respect to D and D'. Hence

P is matrix of id_V with respect to B' and B;
A is the matrix of T with respect to B and D; and
Q^{-1} is the matrix of id_W with respect to D and D'.

Hence by N6.4.2 $Q^{-1}AP$ is the matrix of $\mathrm{id}_W \circ T \circ \mathrm{id}_V$ with respect to B' and D'. Since $T = \mathrm{id}_W \circ T \circ \mathrm{id}_V$ by A.5.2, this gives $A' = Q^{-1}AP$. $\square$

The preceding theorem can be visualized in the following commutative diagram

$$\begin{tikzcd}
& V \arrow[rrrr, "T"] \arrow[dl, "C_{B'}"'] \arrow[dr, "C_B"] & & & & W \arrow[dl, "C_D"'] \arrow[dr, "C_{D'}"] & \\
\mathbb{R}^n \arrow[rr, "L_P"] \arrow[rrrrrr, bend right, "L_{A'}"'] & & \mathbb{R}^n \arrow[rr, "L_A"] & & \mathbb{R}^m \arrow[rr, "L_{Q^{-1}}"] & & \mathbb{R}^m
\end{tikzcd}$$

or if your prefer

$$
\begin{tikzcd}
V \arrow[rrr, "T"] \arrow[dr, "C_B"] \arrow[dd, "C_{B'}"'] & & & W \arrow[dl, "C_D"'] \arrow[dd, "C_{D'}"] \\
& \mathbb{R}^n \arrow[r, "L_A"] & \mathbb{R}^m \arrow[dr, "L_{Q^{-1}}"] & \\
\mathbb{R}^n \arrow[ur, "L_P"] \arrow[rrr, "L_{A'}"'] & & & \mathbb{R}^m
\end{tikzcd}
$$

N6.8 Isomorphism

Corollary 6.30 (Classification Theorem for finite dimensional vector spaces)**.** *Let* $\mathbf{V}$ *be finite dimensional vector space and* $\mathbf{W}$ *a vector space. Then* $\mathbf{V}$ *is isomorphic to* $\mathbf{W}$ *if and only if* $\mathbf{W}$ *is finite dimensional and* $\dim \mathbf{V} = \dim \mathbf{W}$.

Proof. $\Longrightarrow$: Suppose first that $T : \mathbf{V} \to \mathbf{W}$ is an isomorphism and let $B = (v_1, \dots, v_n)$ be basis for V. Then by N6.6.6 $\big(T(v_1), \dots, T(v_n)\big)$ is basis for W. Thus W is finite dimensional and $\dim W = n = \dim V$.

$\Longleftarrow$: Suppose next that $\mathbf{W}$ is finite dimensional and $\dim W = \dim V$. Let $D = (w_1, w_2 \dots, w_n)$ be basis for $\mathbf{W}$. By 6.9 there exists a linear function $T : \mathbf{V} \to \mathbf{W}$ with $T(v_i) = w_i$ for all $1 \leq i \leq n$. Then

$$\big(T(v_1), \dots, T(v_n)\big) = (w_1, \dots, w_n) = D$$

is a basis for $\mathbf{W}$ and so by N6.6.6 T is invertible and so an isomorphism. Hence V is isomorphic to W. $\square$

N6.7 Rank and Nullity

Lemma N6.7.1. *Let* $T : \mathbf{V} \to \mathbf{W}$ *be an isomorphism,* $\mathbf{X}$ *a subspace of* $\mathbf{V}$ *and* $\mathbf{Y}$ *a subspace of* $\mathbf{W}$. *Suppose that for all* $v \in V$,

$$(*) \qquad v \in X \Longleftrightarrow T(v) \in Y$$

Define the function $S : X \to Y$ *by* $S(x) = T(x)$ *for all* $x \in X$. *Then* S *is an isomorphism and so* X *is isomorphic to* Y.

Proof. By assumption, $T(x) \in Y$ for all $x \in X$ and so also $S(x) \in Y$ for all $x \in X$. Thus S is indeed a function from X to Y.

Let $a, b \in X$ with $S(a) = S(b)$. Then also $T(a) = T(b)$ and since T is 1-1, $a = b$. So S is 1-1.

Let $y \in Y$. Since T is onto, $y = T(v)$ for some $v \in V$. Then $T(v) = y \in Y$ and by (*), $v \in X$. Hence $S(v) = T(v) = y$ and S is onto.

We proved that S is 1-1 and onto, and so by Theorem 6.6 S is invertible.

Let $a, b \in X$ and $r, s \in \mathbb{R}$. Then using N6.1.4

$$S(ra + sb) = T(ra + sb) = rT(a) + sT(b) = rS(a) + sS(b),$$

and so S is linear. We proved that S is invertible and linear. So by definition, S is an isomorphism and X is isomorphic to Y. $\square$

Lemma N6.7.2. *Let* $\mathbf{V}$ *and* $\mathbf{W}$ *be finite dimensional vector spaces with basis* B *and* D *respectively. Let* $T : \mathbf{V} \to \mathbf{W}$ *linear and let* A *be the matrix of* T *with respect to* B *and* D.

(a) Let $v \in V$. *Then* $v \in \ker T$ *if and only if* $[v]_B \in \text{Nul}A$. *In particular,* $\ker T$ *is isomorphic to* $\text{Nul}A$.

(b) Let $w \in W$. *Then* $w \in \operatorname{Im} T$ *if and only if* $[w]_D \in \text{Col}A$. *In particular,* $\operatorname{Im} T$ *is isomorphic to* $\text{Col}A$.

Proof. Recall first that by N6.2.8, C_D and L_B are isomorphisms and so are 1-1, linear and onto.

(a) Let $v \in V$. Then

$$\begin{array}{lll} & v \in \ker T & \\ \iff & T(v) = \mathbf{0} & - \text{ definition of } \ker T \\ \iff & [T(v)]_D = \mathbf{0} & - \text{ Since } C_D \text{ is 1-1 and } [\mathbf{0}]_D = \mathbf{0} \\ \iff & A[v]_B = \mathbf{0} & - \text{ since} [T(v)]_D = A[v]_B \text{ by 6.11} \\ \iff & [v]_B \in \text{Nul}A & - \text{ definition of Nul}A \\ \iff & C_B(v) \in \text{Nul}A & - \text{ definition of } C_B \end{array}$$

Thus the first statement in (a) holds. Since C_B is an isomorphism, N6.7.1 shows that $\ker T$ is isomorphic to $\text{Nul}A$.

(b) Let $w \in W$. Then

$$
\begin{array}{lcl}
& w \in \text{Im } T & \\
\iff & w = T(v) \text{ for some } v \in V & - \text{ definition of Im } T \\
\iff & C_D(w) = C_D\big(T(v)\big) \text{ for some } v \in V & - C_D \text{ is 1-1} \\
\iff & C_D(w) = L_A\Big(C_B(v)\Big) \text{ for some } v \in V & - \text{ since } C_D \circ T = L_A \circ C_B \text{ by 6.11} \\
\iff & C_D(w) = L_A(x) \text{ for some } x \in \mathbb{R}^n & - C_B \text{ is onto and so} \\
& & \quad v \in V \text{ if and only } v = C_B(x) \text{ for some } x \in \mathbb{R}^n \\
\iff & C_D(w) \in \text{Im } L_A & - \text{ definition of Im } L_A \\
\iff & C_D(w) \in \text{Col}A & - \text{ since Im } L_A = \text{Col}A \text{ by 6.22} \\
\iff & [w]_D \in \text{Col}A & - \text{ definition of } C_D
\end{array}
$$

Thus the first statement in (b) holds. Since C_D is an isomorphism, N6.7.1 shows that Im T is isomorphic to Col$A$. □

Theorem 6.27 (Dimension Theorem)**.** *Let $T : \mathbf{V} \to \mathbf{W}$ be a linear. If $\mathbf{V}$ and $\mathbf{W}$ are finite dimensional, then*

$$\dim \ker T + \dim \text{Im } T = \dim V$$

Proof. Let B and D be bases for $\mathbf{V}$ and $\mathbf{W}$, respectively. Let A be matrix of T with respect to B and D. Let $n = \dim V$. By Theorem 6.25

$$(*) \qquad \dim \text{Nul}A + \dim \text{Col}A = n = \dim V$$

By N6.7.2 $\ker T$ is isomorphic to Nul$A$ and Im T is isomorphic to Col$A$. Thus by 6.30

$$\dim \text{Nul}A = \dim \ker T \text{ and } \dim \text{Col}A = \dim \text{Im } T.$$

Together with (*) this proves the theorem. □

Chapter N7

Determinants

N7.2 Definition and Properties

Definition N7.2.1. *Let n be non-negative integer and $\alpha \in \mathbb{R}$. A function*

$$\mathrm{D} : \mathbb{M}(n, n) \to \mathbb{R}$$

is called an α-based determinant function provided that if the following three statements hold:

(i) Let A be an $n \times n$ matrix and $1 \leq j \leq n$. Then the function

$$\mathrm{D}_{A_j} : \mathbb{R}^n \to \mathbb{R}, \text{ with } \mathrm{D}_{A_j}(x) = \mathrm{D}\big(C_j x A\big) \text{ for all } x \in \mathbb{R}^n$$

is linear. (So D_{A_j} is the function obtained from D by keeping all Columns but Column j constant.)

(ii) Let A be an $n \times n$ matrix and $1 \leq i < j \leq n$. If $a_i = a_j$, then $\det A = 0$

(iii) $\mathrm{D}(I) = \alpha$.

A regular determinant function is a 1-based determinant function.

In the following we will show that for each $n \in \mathbb{N}$ and $\alpha \in \mathbb{R}$ there exists a unique α-based determinant function from $\mathbb{M}(n, n)$ to $\mathbb{R}$.

Lemma N7.2.2. *Let n be positive integer, $\alpha \in \mathbb{R}$, $\mathrm{D} : \mathbb{M}(n, n) \to \mathbb{R}$ an α-based determinant function, and $A \in \mathbb{M}(n, n)$.*

(a) Let $1 \leq j \leq n$ and $r \in \mathbb{R}$. Let B be the matrix obtain by multiplying Column j of A with r. Then $\mathrm{D}(B) = r\mathrm{D}(A)$.

(b) Let $1 \leq j, k \leq n$ with $j \neq k$. Let B be the matrix obtain adding r-times column k of A to Column j of A. Then $\mathrm{D}(B) = \mathrm{D}(A)$.

(c) Let $1 \leq j, k \leq n$ with $j \neq k$. Let B be the matrix obtain from A by interchanging Column j and Column k of A. Then $\mathrm{D}(B) = -\mathrm{D}(A)$

(d) If A has a zero column, then $\mathrm{D}(A) = 0$.

Proof. We will just write $\langle x \rangle$ for $C_j x A$. $A = \langle a_j \rangle$ and $\mathrm{D}_{A_j}(x) = \mathrm{D}(\langle x \rangle)$.

(a) Note that $B = \langle r a_j \rangle$. Since D_{A_j} is linear,

$$\mathrm{D}(B) = \mathrm{D}\big(\langle r a_j \rangle\big) = \mathrm{D}_{A_j}(r a_j) = r\mathrm{D}_{A_j}(a_j) = r\mathrm{D}\big(\langle a_j \rangle\big) = \mathrm{D}(A).$$

(b) Note that $B = \langle a_j + r a_k \rangle$. Since D_{A_j} is linear,

$$\begin{aligned} \mathrm{D}(B) = \mathrm{D}\big(\langle a_j + r a_k \rangle\big) &= \quad \mathrm{D}_{A_j}(a_j + r a_k) &&= \mathrm{D}_{A_j}(a_j) + r\mathrm{D}_{A_j}(a_k) \\ &= \mathrm{D}\big(\langle a_j \rangle\big) + r\mathrm{D}\big(\langle a_k \rangle\big) &&= \mathrm{D}(A) + r\mathrm{D}\big(\langle a_k \rangle\big). \end{aligned}$$

Note that Columns j and k of $\langle a_k \rangle$ are both equal to a_k, and so Condition (ii) of a determinant function shows that shows that $\mathrm{D}\big(\langle a_k \rangle\big) = 0$. So $\mathrm{D}(B) = \mathrm{D}(A)$.

(c) We will just write $\langle x, y \rangle$ for $C_{jk} x y A$. Then $A = \langle a_j, a_k \rangle$ and $B = \langle a_k, a_j \rangle$. We will show how to obtain B from A via a sequence of column operation as in (a) and (b)

$$\begin{array}{rcl c} & & & A = \langle a_j, a_k \rangle \\ 1C_k + C_j & \to & C_j & \langle a_k + a_j, a_k \rangle \\ (-1)C_j + C_k & \to & C_k & \langle a_k + a_j, (-1)(a_k + a_j) + a_k \rangle = \langle a_k + a_j, -a_j \rangle \\ 1C_k + C_j & \to & C_j & \langle (-a_j) + (a_k + a_j), -a_j \rangle = \langle a_k, -a_j \rangle \\ (-1)C_k & \to & C_k & \langle a_k, a_j \rangle \end{array}$$

Note that all but the last operation are as in (b) and so do not change the determinant. The last one multiplies the determinant by -1. So $\mathrm{D}(B) = -\mathrm{D}(A)$.

(d) Suppose column j is zero. Since D_{A_j} is linear, Theorem 6.2 gives $\mathrm{D}(A) = \mathrm{D}_{A_j}(a_j) = \mathrm{D}_{A_j}(\mathbf{0}) = 0$. □

Algorithm N7.2.3. *Let A be an $n \times n$ matrix and D an α-based determinant function. Let B the reduced column echelon form of A and $E_1, \ldots, E_l$ a sequence of elementary column operation which transforms A into B. If E_i is $rC_j \to C_j$, put $r_i = \frac{1}{r}$. If E_i is $rC_k + C_j + \to C_j$, put $r_i = 1$ and if E_i is $C_j \leftrightarrow C_k$, put $r_i = -1$.*

(a) If $B = I$, then $\mathrm{D}(A) = r_1 r_2 \ldots r_l \alpha$.

(b) If $B \neq I$, then $\mathrm{D}(A) = 0$.

Proof. Inductively define

$$A_0 = A, A_1 = E_1(A_0), A_2 = E_2(A_1), \ldots, A_l = E_l(A_{l-1})$$

So A_i is the matrix obtained from A_{i-1} via the elementary column operation E_i. Thus $A_l = B$. Then by N7.2.2 $\mathrm{D}(A_i) = \frac{1}{r_i}\mathrm{D}(A_{i-1})$. Thus $\mathrm{D}(A_{i-1}) = r_i\mathrm{D}(A_1)$ and so

$$\mathrm{D}(A_0) = r_1\mathrm{D}(A_1) = r_1r_2\mathrm{D}(A_2) = r_1r_2r_3\mathrm{D}(A_3) = \ldots = r_1r_2\ldots r_l\mathrm{D}(A_l).$$

Hence

$$(*) \qquad \mathrm{D}(A) = r_1\ldots r_l\mathrm{D}(B)$$

Suppose now that $B = I$. By definition $\mathrm{D}(I) = \alpha$ and so $\mathrm{D}(A) = r_1\ldots r_l\alpha$ by (*). Thus (a) holds.

Suppose next that $B \neq I_n$. Then by the column version of N5.2.3 B has a zero column. Thus $\mathrm{D}(B) = 0$ by N7.2.2 and so $\mathrm{D}(A) = 0$ by (*). Thus (b) holds. □

Corollary N7.2.4. *Let $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}$. Then there exists at most one α-based determinant function from $\mathbb{M}(n,n)$ to $\mathbb{R}$.*

Proof. Let D an α-based determinant function. Then N7.2.3 tells us how to compute $\mathrm{D}(A)$ and so D is unique. □

Corollary N7.2.5. *Let $\alpha \in \mathbb{R}$ and $n \in \mathbb{N}$. Suppose* $\det : \mathbb{M}(n,n) \to \mathbb{R}$ *is a regular determinant function. Then* $\alpha\det$ *is the unique α-based determinant function.*

Proof. Let A be an $n \times n$ matrix, $1 \leq j \leq n$ and $x \in \mathbb{R}^n$. Then

$$\begin{aligned}
(\alpha\det)_{A_j}(x) &= (\alpha\det)(C_jxA) && \text{– definition of } (\alpha\det)_{A_j} \\
&= \alpha\big(\det(C_jxA) && \text{– definition of multiplication for functions} \\
&= \alpha\big(\det{}_{A_j}(x)\big) && \text{– definition of } \det{}_{A_j} \\
&= (\alpha\det{}_{A_j})(x) && \text{– definition of multiplication for functions}
\end{aligned}$$

Thus $(\alpha\det)_{A_j} = \alpha\det_{A_j}$ by A.2.2. Since $\det_{A_j}$ is linear, also $(\alpha\det)_{A_j}$ is linear by N6.2.4(c). If $a_j \neq a_k$ for distinct j,k, then $\det A = 0$ and so also $(\alpha\det)(A) = \alpha(\det A) = 0$. Finally $(\alpha\det)(I) = \alpha(\det(I)) = \alpha 1 = \alpha$ and so $\alpha\det$ is a α-based determinant function. By N7.2.4, $\alpha\det$ is the unique such function. □

Corollary N7.2.6. *Let A be an $n\times n$ matrix and* det *a regular determinant function. Then the following statements are equivalent:*

(a) $\det(A) \neq 0$.

(b) The reduced column-echelon form of A is I.

(c) A is invertible.

Proof. (a) $\iff$ (b) : Let B and $r_1, \ldots, r_l$ be as in N7.2.3. Then $r_i \neq 0$ for all i and so also $r_1 \ldots r_l \neq 0$. So if $B = I$, then $\det A = r_1 r_2 \ldots r_l \neq 0$ and if the reduced $B \neq I$, then $\det A = 0$.

(b) $\iff$ (c) : This is the column version of N5.2.4(a). □

Lemma N7.2.7. *Let A be an $m \times n$-matrix and B an $n \times p$ matrix. Let $x \in \mathbb{R}^p$ and $1 \leq k \leq p$. Then $C_k(Ax)(AB) = A(C_k x B)$.*

Proof. Let $D = C_k x B$ and $1 \leq l \leq p$. We need to show that

(*) Column l of AD is equal to Column l of $A(C_k x B)$.

Column k of D is x and so Column k of AD is Ax. Column k of $C_k(Ax)(AB)$ is also equal to Ax and so (*) holds for $l = k$.

If $l \neq k$, then column l of D is b_l and so Column l of AD is Ab_l. Column l of AB is also equal to Ab_l and so (*) also holds for $l \neq k$. □

Theorem 7.7. *Let* $\det : \mathbb{M}(n,n) \to \mathbb{R}$ *be a regular determinant function and $A, B \in \mathbb{M}(n,n)$. Then* $\det(AB) = \det(A)\det(B)$.

Proof. Fix $A \in \mathbb{M}(n,n)$. Put $\alpha = \det(A)$ and define $\mathrm{D} : \mathbb{M}(n,n) \to \mathbb{R}$ by $\mathrm{D}(B) = \det(AB)$. We will first show that D is an α-based determinant function. Let $1 \leq j \leq n$, and $x \in \mathbb{R}^n$.

$$\begin{aligned}
\mathrm{D}_{B_j}(x) &= \mathrm{D}(C_j x B) && \text{– definition of } \mathrm{D}_{B_j} \\
&= \det\Big(A(C_j x B)\Big) && \text{– definition of D} \\
&= \det\Big(C_j(Ax)(AB)\Big) && \text{– N7.2.7} \\
&= \det\nolimits_{(AB)_j}(Ax) && \text{– definition of } \det\nolimits_{(AB)_j} \\
&= \det\nolimits_{(AB)_j}\big(L_A(x)\big) && \text{– definition of } L_A \\
&= \Big(\det\nolimits_{(AB)_j} \circ L_A\Big)(x) && \text{– definition of composition}
\end{aligned}$$

Hence $\mathrm{D}_{B_j} = \det_{(AB)_j} \circ L_A$. By definition of a determinant function, $\det_{(AB)_j}$ is linear. By N6.1.7 L_A is linear and so by 6.7 also D_{B_j}. linear. So D fulfills (i) in the definition of an α-based determinant function.

Suppose that columns j and k of B are both equal to some $x \in \mathbb{R}^n$. Then columns j and k of AB are both equal to Ax. Thus $\det(AB) = 0$ and so $\mathrm{D}(B) = 0$. Hence condition (ii) is also fulfilled. Now $\mathrm{D}(I) = \det(AI) = \det(A) = \alpha$ and so D is an α-based determinant function. Thus by N7.2.5 $\mathrm{D} = \alpha \det$. Hence

$$\det(AB) = \mathrm{D}(B) = (\alpha \det)(B) = \alpha \det(B) = \det(A)\det(B).$$

□

The preceding theorem tells us how linear transformation effect volume. The volume of the box spanned by vectors $b_1, \ldots, b_n$ in $\mathbb{R}^n$ is $\det B$ where B is the matrix $[b_1, \ldots, b_n]$. Under the linear transformation L_A, this box is mapped to the box spanned by $Ab_1, \ldots, Ab_n$. This box has volume $\det[Ab_1, \ldots, Ab_n] = \det AB = \det A \det B$. Approximating the volume of an arbitrary region in $\mathbb{R}^n$ by decomposing the region into small boxes we conclude that the volume of the image of region under L_A is $\det A$ times the volume of the original region.

N7.3 Existence

Lemma N7.3.1. *Let A be $n \times n$ matrix and D a determinant function. Let $1 \le i \le j$ and define the matrix B by*

$$b_k = \begin{cases} a_k & \text{if } 1 \le k < i \\ a_j & \text{if } k = i \\ a_{k-1} & \text{if } i < k \le j \\ a_k & \text{if } j < k \le n \end{cases},$$

that is

$$B = [a_1, \ldots, a_{i-1}, a_j, a_i, a_{i+1}, \ldots, a_{j-2}, a_{j-1}, a_{j+1}, \ldots a_n]$$

Then $\mathrm{D}(B) = (-1)^{j-i}\,\mathrm{D}(A)$.

Proof. Observe that B can be transformed into A by the following sequence of $j - i$ elementary row operations:

$$C_i \leftrightarrow C_{i+1}, \quad C_{i+1} \leftrightarrow C_{i+2}, \quad \ldots, \quad C_{j-1} \leftrightarrow C_j$$

Each of these operation multiplies the determined by -1 and so $\mathrm{D}(B) = (-1)^{j-i}\mathrm{D}(A)$. □

Lemma N7.3.2. *Let n be a positive integer and $1 \le i \le n$.*

(a) Define $\sigma_i : \mathbb{R}^n \to \mathbb{R}$ by $\sigma_i(a_1, \ldots, a_n) = a_i$. Then σ_i is linear.

(b) Define $\pi_i : \mathbb{R}^n \to \mathbb{R}^{n-1}$ by $\pi_i(a_1, \ldots, a_n) = (a_1, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n)$. Then π_i is linear.

(c) Define $\tau_i : \mathbb{R}^{n-1} \to \mathbb{R}^n$ by $\tau_i(a_1, \ldots, a_{n-1}) = (a_1, \ldots, a_{i-1}, 0, a_i, \ldots, a_{n-1})$. Then τ_i is linear.

Proof. (a) This can be proved by direction computation or by observing that $\sigma_i = L_{e_i}$.

(b) This can be proved by direction computation or by observing that $\pi_i = L_B$ where B is the list $(e_1, \ldots, e_{i-1}, 0, e_i, \ldots, e_{n-1})$ in $\mathbb{R}^{n-1}$.

(c) This can be proved by direction computation or by observing that $\tau_i = L_B$, where B is the list $(e_1, \ldots, e_{i-1}, e_{i+1}, \ldots, e_n)$ in $\mathbb{R}^n$.

□

Theorem N7.3.3. *Let n be a non-negative integer.*

(a) There exists a unique regular determinant function $\det : \mathbb{M}(n,n) \to \mathbb{R}$.

(b) Let $1 \leq i \leq n$ and A an $n \times n$-matrix. Then

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

where A_{ij} is the $(n-1) \times (n-1)$ matrix obtained from A by deleting row i and column j.

Proof. In view of N7.2.4 we only need to show the existence of a regular determinant function.

The proof is induction on n. For $n = 0$, $\mathbb{M}(n,n)$ has a unique element namely the empty matrix $I_0 = []$. If we define $\det([]) = 1$, then det is a regular determinant function.

So suppose $n > 0$ and that the theorem holds for $n-1$ in place of n. Fix $1 \leq i \leq n$ and define a function $\det : \mathbb{M}(n,n) \to \mathbb{R}$ by

$$(*) \qquad \det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$$

for all $A \in \mathbb{M}(n,n)$. We need to verified that det is a regular determinant function. Fix $1 \leq k \leq n$. We will show that $\det_{A_k}$ is linear. For $1 \leq j < k$ define $F_j = \det_{(A_{ij})_{k-1}}$. For $k < j \leq n$ define $F_j = \det_{(A_{ij})_k}$. Let $x \in \mathbb{R}^n$ and put $x' = \pi_i(x)$. (So x' is obtained from x by deleting entry i.) If $j < k$, then $(C_k x A)_{ij} = C_{k-1} x' A_{ij}$ and so $\det(C_k x A)_{ij} = F_j(x')$. If $k = j$, then $(C_k x A)_{ij} = A_{ij}$ so $\det(C_k x A)_{ij} = \det A_{ij}$. If $k < j \leq n$, then $(C_k x A)_{ij} = C_k x' A_{ij}$ and so $\det(C_k x A)_{ij}) = F_j(x')$. Let $\tilde{a}_{ij}$ be the (i,j)-entry of $C_k x A)$. So $\tilde{a}_{ij} = a_{ij}$ if $j \neq k$ and $\tilde{a}_{ik} = x_k$. Thus

$$\begin{aligned}
\det_{A_k}(x) = \det(C_k x A) &= \textstyle\sum_{j=1}^{n} (-1)^{i+j} \tilde{a}_{ij} \det(C_k x A)_{ij} \\
&= \textstyle\sum_{j=1}^{k-1} a_{ij} F_j(x') + x_k \det(A_{ij}) + \sum_{j=k+1}^{n} a_{ij} F_j(x') \\
&= \textstyle\sum_{j=1}^{k-1} a_{ij} F_j(\pi_j(x)) + \det(A_{ij}) \sigma_i(x) + \sum_{j=k+1}^{n} a_{ij} F_j(\pi_i(x)))
\end{aligned}$$

and so

$$\det_{A_k} = \sum_{j=1}^{k-1} a_{ij} F_j \circ \pi_i + \det(A_{ij}) \sigma_i + \sum_{j=k+1}^{n} a_{ij} F_j \circ \pi_j$$

By definition of a determinant function F_j is linear. By N7.3.2 π_i and σ_i are linear. So also $F_j \circ \pi_i$ is linear. Thus $\det_{A_k}$ is a linear combination of linear functions and so is linear by N6.2.4. Thus Condition (i) in n the definition of a determinant function holds.

The (i,j) entry of I_n is zero for $j \neq i$ and 1 for $i = j$. Also $(I_n)_{ii} = I_{n-1}$ and so $\det(I_n) = (-1)^{i+i} \cdot 1 \cdot \det(I_{n-1}) = 1 \cdot 1 = 1$. Thus also Condition (iii) in the definition of a determinant function holds.

If remains to show that $\det(A) = 0$ if A has two equal columns So suppose that columns r and s of A are equal for some $1 \leq r < s \leq n$. If $j \neq r$ and $j \neq s$, then A_{ij} has two equal columns and so $\det A_{ij} = 0$. Thus

$$(*) \qquad \begin{aligned} \det A &= (-1)^{i+r} a_{ir} \det A_{ir} + (-1)^{i+s} a_{is} \det A_{is} \\ &= (-1)^{i+r} a_{ir} \Big(\det A_{ir} + (-1)^{s-r} \det A_{is} \Big) \end{aligned}$$

We have

$$\begin{aligned} A_{ir} &= [\ a'_1,\ \ldots,\ a'_{r-1},\ a'_{r+1},\ a'_{r+1},\ \ldots,\ a'_{s-1},\ \ a'_s,\ \ a'_{s+1},\ \ldots,\ a'_n\ ] \\ A_{is} &= [\ a'_1,\ \ldots,\ a'_{r-1},\ \ a'_r,\ \ a'_{r+1},\ \ldots,\ a'_{s-2},\ a'_{s-1},\ a'_{s+1},\ \ldots,\ a'_n\ ] \end{aligned}$$

Since $a'_s = a'_r$ we conclude from N7.3.1 applied with $i = r$ and $j = s-1$, that $\det(A_{ir}) = (-1)^{s-r-1} \det(A_{is}) = -(-1)^{s-r} \det(A_{is})$ and so (*) shows that $\det A = 0$. So also Condition 2 in the definition of a determinant function holds and so det is a regular determinant function.

□

N7.4 Cramer's Rule

Definition N7.4.1. *Let A be an $n \times n$ matrix and* $\det : \mathbb{M}(n,n) \to \mathbb{R}$ *a regular determinant function. Then the adjoint of A is the $n \times n$ matrix B with $b_{ij} = \det(C_i e_j A)$.*

Theorem N7.4.2. *A be an $n \times n$ matrix, and* $\det : \mathbb{M}(n,n) \to \mathbb{R}$ *a regular determinant function and B the adjoint of A.*

(a) Let $x \in \mathbb{R}^n$ and $1 \leq i \leq n$. Then $\det(C_i x A) = b_{i1}x_1 + \ldots + b_{in}x_n = b^i x$.

(b) Let $x \in \mathbb{R}^n$. Then $Bx = \Big(\det(C_i x A)\Big)_{i=1}^n = \Big(\det(C_1 x A), \ldots, \det(C_n x A)\Big)$.

(c) Let $1 \leq i, j \leq n$. Then $b^i a_j = 0$ if $i \neq j$ and $b^i a_j = \det(A)$ if $i = j$.

(d) $BA = \det(A) I$. In particular, if $\det(A) \neq 0$, A is invertible and $A^{-1} = \frac{1}{\det(A)} B$.

(e) Let $x, y \in \mathbb{R}^n$ with $Ax = y$. Then $(\det A)x = By = \Big(\det(C_i y A)\Big)_{i=1}^n$. In particular, if $\det A \neq 0$,

$$x = \frac{1}{\det(A)} By = \left(\frac{\det(C_i y A)}{\det(A)} \right)_{i=1}^n .$$

Proof. (a) We have $\det_{A_i}(e_j) = \det(C_i e_j A) = b_{ij} = b_{ij}1$. So b^i is the matrix of the linear function $\det_{A_i}$ with respect to the basis $(e_1, \ldots, e_n)$ for $\mathbb{R}^n$ and the basis (1) for $\mathbb{R}$. Thus (a) follows from 6.11.

(b) Since entry i of Bx is $b^i x$, (b) follows from (a).

(c) By (a) $b^i a_j = \det(C_i a_j A)$. If $i \neq j$, then columns i and j of $C_i a_j A$ are both equal to a_j and so $b^i a_j = \det(C_i a_j A) = 0$. If $i = j$, then $C_i a_j A = A$ and so $b^i a_j = \det(A)$.

(d) follows from (c) and $BA = (b^i a_j)_{i=1,j=1}^{n,n}$.

(e) From $Ax = y$ we get

$$By = B(Ax) = (BA)x = \big((\det A)I\big)x = (\det A)(Ix) = (\det A)x.$$

The second equality now follows from (b). $\square$

Geometrically, (c) means that the vector b^i is perpendicular to the subspace of $\mathbb{R}^n$ spanned by $(a_1, \ldots, a_{i-1}, a_{i+1}, \ldots a_n)$ and that $\det(A)$ can be computed via the dot product of b^i and a_i.

Theorem N7.4.3. *Let* $\det : \mathbb{M}(n,n) \to \mathbb{R}$ *a regular determinant function. Let A be an $n \times n$ matrix, $1 \leq i, j \leq n$ and let A_{ij} the $(n-1) \times (n-1)$-matrix obtained from A by deleting Row i and Column j of A.*

(a) $\det(C_j e_i A) = (-1)^{i+j} \det(A_{ij})$.

(b) Let B be the adjoint of A. Then $B = \Big((-1)^{i+j} \det(A_{ji})\Big)_{i=1,j=1}^{n,n}$.

(c) $\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A_{ij})$.

Proof. (a) For $1 \leq k \leq n-1$ define $\hat{k} = k$ if $k < j$ and $\hat{k} = k+1$ if $j \leq k \leq n-1$. For $D \in M(n-1, n-1, \mathbb{R})$ let D^* be the $n \times n$ matrix defined by

$$d_k^* = \begin{cases} \tau_i(d_k) & \text{if } 1 \leq k < j \\ e_i & \text{if } k = j \\ \tau_i(d_{k-1}) & \text{if } j < k \end{cases}$$

Then $D^*_{ij} = D$ and $d_{\hat{k}} = \tau_i(d_k)$ for all $1 \leq k \leq n-1$. Put $\alpha = \det(I^*_{n-1})$.

We claim the function $\mathrm{E} : \mathbb{M}(n-1, n-1) \to \mathbb{R}$ defined by $\mathrm{E}(D) = \det(D^*)$ is an α-based determinant function. Note that

$$\mathrm{E}_{D_k}(x) = \det\Big((C_k x D)^*\Big) = \det\Big(C_{\hat{k}} \tau_i(x) D^*\Big) = \det_{D^*_{\hat{k}}}\big(\tau_i(x)\big)$$

and thus $\mathrm{E}_{D_k} = \det_{D^*_{\hat{k}}} \circ \tau_i$. Since both $\det_{D^*_{\hat{k}}}$ and τ_i are linear, E_{D_k} is linear. If columns k and l of D are both equal to $x \in \mathbb{R}^n$, then columns $\hat{k}$ and $\hat{l}$ are both equal to $\tau_i(x)$. So $\mathrm{E}(D) = \det(D^*) = 0$. By definition of α, $\mathrm{E}(I_{n-1}) = \alpha$ and so E is indeed an α-based determinant function. Hence $E = \alpha \det$.

Note that $\tau_i(e_k) = e_k$ for $1 \le k < i$ and $\tau_i(e_k) = e_{k+1}$ for $i \le k \le n-1$. This allows us to compute I^*_{n-1}.

For $i < j$,

$$\begin{aligned} I^*_{n-1} &= [\ e_1,\ \dots,\ e_{i-1},\ e_{i+1},\ e_{i+2},\ \dots,\ \ e_j,\ \ e_i,\ e_{j+1},\ \dots,\ e_n\ ], \\ I_n &= [\ e_1,\ \dots,\ e_{i-1},\ \ e_i,\ \ e_{i+1},\ \dots,\ e_{j-1},\ e_j,\ e_{j+1},\ \dots,\ e_n\ \ ] \end{aligned}$$

and so by N7.3.1 applied with (i, j, I^*_{n-1}, I_n) in place of (i, j, A, B), $\det I_n = (-1)^{j-i} \det I^*_{n-1}$ and so $\alpha = (-1)^{i+j}$.

For $i = j$,

$$I^*_{n-1} = [e_1, \dots, e_{i-1}, e_i, e_{i+1}, \dots, e_n] = I_n$$

and so $\alpha = 1 = (-1)^{i+j}$.

For $i > j$,

$$\begin{aligned} I_n &= [\ e_1,\ \dots,\ e_{j-1},\ e_j,\ e_{j+1},\ \dots,\ e_{i-1},\ \ e_i,\ \ e_{i+1},\ \dots,\ e_n\ ], \\ I^*_{n-1} &= [\ e_1,\ \dots,\ e_{j-1},\ e_i,\ \ e_j,\ \ \dots,\ e_{i-2},\ e_{i-1},\ e_{i+1},\ \dots,\ e_n\ \ ] \end{aligned}$$

and so by N7.3.1, applied with (j, i, I_n, I^*_{n-1}) in place of (i, j, A, B), $\det(I^*_{n-1}) = (-1)^{i-j} \det I$ and $\alpha = (-1)^{i+j}$.

Hence in all cases $\alpha = (-1)^{i+j}$ and so $\mathrm{E} = (-1)^{i+j} \det$.

Let C be the matrix obtained from $C_j e_i A$ by adding (for each $1 \le k \le n$ with $k \ne j$) $-a_{ik}$ times column j to column k. Then $C = A^*_{ij}$ and

$$\det(C_j e_i A) = \det C = \det A^*_{ij} = \mathrm{E}(A_{ij}) = (-1)^{i+j} \det A_{ij}$$

and so (a) holds.

(b) Let B be the adjoint of A. Then $b_{ji} = \det(C_j e_i A) = (-1)^{i+j} \det A_{ij}$ and so (b) holds.

(c) Using N7.4.2(a) and (a) we compute

$$\det A = b^j a_j = \sum_{i=1}^n b_{ji} a_{ij} = \sum_{i=1}^n a_{ij} b_{ji} = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij})$$

□

Chapter N8

Eigenvalues and Eigenvectors

N8.1 Definitions

Definition 8.1. *Let $T : \mathbf{V} \to \mathbf{V}$ be linear and $\lambda \in \mathbb{R}$.*

(a) $v \in V$ is called an eigenvector of T associated to λ if $v \neq \mathbf{0}$ and $T(v) = \lambda v$.

(b) λ is called an eigenvalue of T if there exists an eigenvector of T associated to λ.

(c) $E_T(\lambda) = \{v \in V \mid T(v) = \lambda v\}$. (So $E_T(\lambda)$ consists of the eigenvectors of T associated to λ and the zero vector.) $E_T(\lambda)$ is called the eigenspace of T associated to λ

We will use the same terminology if V is replaced by $\mathbb{R}^n$ and T by an $n \times n$ matrix A. So an eigenvalue for A is the same as an eigenvalue for L_A.

Theorem 8.2. *Let $T : \mathbf{V} \to \mathbf{V}$ be linear, let A be an $n \times n$ matrix and let $\lambda \in \mathbb{R}$.*

(a) $E_T(\lambda) = \ker(\lambda \mathrm{id}_V - T)$. In particular, $E_T(\lambda)$ is a subspace of $\mathbf{V}$.

(b) $E_A(\lambda) = \mathrm{Nul}(\lambda I - A)$. In particular, $E_A(\lambda)$ is a subspace of $\mathbb{R}^n$.

(c) λ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$.

Proof. (a) Let $v \in V$. Then

$$\begin{array}{rcl}
 & v \in E_T(\lambda) & \\
\iff & T(v) = \lambda v & - \text{ definition of } E_T(\lambda) \\
\iff & T(v) = \lambda\big(\mathrm{id}_V(v)\big) & - \text{ definition of } \mathrm{id}_V \\
\iff & T(v) = (\lambda \mathrm{id}_V)(v) & - \text{ definition of multiplication for functions} \\
\iff & (\lambda \mathrm{id}_V)(v) - T(v) = \mathbf{0} & - \text{ 1.7(p)} \\
\iff & (\lambda \mathrm{id}_V - T)(v) = \mathbf{0} & - \text{ N6.4.1(a)} \\
\iff & v \in \ker(\lambda \mathrm{id}_V - T) & - \text{ definition of } \ker(\lambda \mathrm{id}_V - T)
\end{array}$$

(b) By N6.4.2 the matrix for $\lambda\mathrm{id}_V - L_A$ with respect to standard basis for $\mathbb{R}^n$ is $\lambda I - A$. Hence by Theorem 6.22 $\ker \lambda\mathrm{id}_V - L_A = \mathrm{Nul}\lambda I - A$. Thus using (a), $E_A(\lambda) = E_{L_A}(\lambda) = \ker(\lambda\mathrm{id}_V - T) = \mathrm{Nul}(\lambda I - A)$

(c) We have

$$
\begin{array}{clr}
 & \lambda \text{ is an eigenvalue of } A & \\
\iff & \text{there exists an eigenvector of } T \text{ associated to } \lambda & - \text{ definition of eigenvalue} \\
\iff & E_A(\lambda) \neq \{\mathbf{0}\} & - \text{ definition of } E_A(\lambda) \\
\iff & \mathrm{Nul}(\lambda I - A) \neq \{\mathbf{0}\} & -\text{(b)} \\
\iff & \lambda I - A \text{ is not invertible} & - \text{ N5.2.1} \\
\iff & \det(\lambda I - A) = 0 & - \text{ N7.2.6}
\end{array}
$$

□

Definition 8.3. *Let A be an $n \times n$-matrix. Then the function $\chi_A : \mathbb{R} \to \mathbb{R}$ defined by*

$$\chi_A(\lambda) = \det(\lambda I - A)$$

for all $\lambda \in \mathbb{R}$ is called the characteristic polynomial of A.

Note that λ is an eigenvalue of A if and only if $\det(\lambda I - A) = 0$ and so if and only if $\chi_A(\lambda) = 0$, that is if and only λ is a root of χ_A.

Theorem N8.1.4. *Let $T : \mathrm{V} \to \mathrm{V}$ be linear and $\lambda_1, \ldots, \lambda_k$ be distinct eigenvalues of T. For $1 \leq i \leq k$ let v_i be an eigenvector of T associated to λ_i. Then $(v_1, \ldots, v_k)$ is linearly independent.*

Proof. The proof is by induction on k. Since the empty list is linearly independent, the theorem holds for $k = 0$. Suppose it holds for $k - 1$. We will show that its also holds for k. For this let $(r_1, \ldots, r_k) \in \mathbb{R}^k$ with

$$(*) \qquad r_1 v_1 + \ldots + r_k v_k = \mathbf{0}.$$

Applying T to both sides and using Theorem 6.2 we get

$$r_1 T(v_1) + \ldots + r_k T(v_k) = \mathbf{0}.$$

Since v_i is an eigenvector associated to λ_i, $T(v_i) = \lambda_i v_i$ and so

$$(**) \qquad r_1 \lambda_1 v_1 + \ldots + r_k \lambda_k v_k = \mathbf{0}.$$

Multiplying equation (*) with λ_k and subtracting from (**) we obtain

$$r_1(\lambda_1-\lambda_k)v_{k-1}+\ldots+r_{k-1}(\lambda_{k-1}-\lambda_k)v_{k-1}=\mathbf{0}.$$

By the induction assumption $(v_1,\ldots,v_{k-1})$ is linearly independent. Hence $r_i(\lambda_i-\lambda_k)=0$ for all $1\leq i<k$. Since $\lambda_i\neq\lambda_k$ for all $1\leq i<k$ we have $\lambda_i-\lambda_k\neq 0$ and so $r_i=0$ for all $1\leq i<k$. Thus (*) implies $r_kv_k=\mathbf{0}$. Since $v_k\neq\mathbf{0}$ this means $r_k=0$ and so $(v_1,\ldots,v_k)$ is linearly independent.

The theorem now follows from the Principal of induction. $\square$

Theorem 8.14. *Let $T:\mathbf{V}\to\mathbf{V}$ be linear and $\lambda_1,\ldots,\lambda_k$ be distinct eigenvalues of T. For each $1\leq i\leq k$ let $(u_{i1},\ldots,u_{il_i})$ be a linearly independent list of eigenvectors of T associated to λ_i. Then*

$$(u_{11},\ldots,u_{1l_1},u_{21},\ldots,u_{2l_2},\ldots\ldots,u_{k1},\ldots,u_{kl_k})$$

is linearly independent.

Proof. Let $r_{ij}\in\mathbb{R}$ for $1\leq i\leq k$ and $1\leq j\leq l_k$ with

$$r_{11}u_{11}+\ldots+r_{1l_1}u_{1l_1}+\ldots\ldots+r_{k1}u_{k1}+\ldots+r_{kl_k}u_{kl_k}=\mathbf{0}.$$

For $1\leq i\leq k$, put $v_i=r_{i1}u_{i1}+\ldots+r_{il_i}u_{il_i}$. By 8.2 $E_T(\lambda_i)$ is a subspace of $\mathbf{V}$ and so $v_i\in E_T(\lambda_i)$. Thus $v_i=\mathbf{0}$ or v_i is an eigenvector of T associated to λ_i. Also

$$v_1+\ldots+v_k=r_{11}u_{11}+\ldots+r_{1l_1}u_{1l_1}+\ldots+r_{k1}u_{k1}\ldots+r_{kl_k}u_{kl_k}=\mathbf{0}.$$

Let $(w_1,\ldots,w_l)$ be the sublist of $(v_1,\ldots,v_k)$ consisting of the non-zero $v_i's$. Then

$$1w_1+\ldots+1w_l=w_1+\ldots+w_l=v_1+\ldots+v_k=\mathbf{0}.$$

If $l\neq 0$ this contradicts N8.1.4. Thus $l=0$ and so $v_i=0$ for all $1\leq i\leq k$. Hence

$$r_{i1}u_{i1}+\ldots+r_{il_i}u_{il_i}=v_i=\mathbf{0}$$

for all $1\leq i\leq k$ and since $(u_{i1},\ldots,u_{il_i})$ is linearly independent we conclude that $r_{ij}=0$ for all $1\leq j\leq l_i$. Thus $(u_{11},\ldots,u_{1l_1},u_{21},\ldots,u_{2l_2},\ldots\ldots,u_{k1},\ldots,u_{kl_k})$ is indeed linearly independent. $\square$

N8.2 Similarity

Definition 8.4. *Let A and A' be $n\times n$ matrices. We say that A is similar to A' and write $A\sim A'$ if there exists an invertible $n\times n$-matrix P with $A'=P^{-1}AP$.*

Lemma N8.2.2. *Let $\mathbf{V}$ be an n-dimensional vector space, $T:\mathbf{V}\to\mathbf{V}$ linear, B a basis for $\mathbf{V}$ and A the matrix of T with respect to B. Let A' be a $n\times n$ matrix. Then A' is similar to A if and only if there exists a basis B' of $\mathbf{V}$ such that A' is the matrix of T with respect B'.*

Proof. Suppose first A' is the matrix of T with respect to some basis B' of $\mathbf{V}$. Let P be the change-of-basis matrix from B' to B. Then by 6.18 $A' = P^{-1}AP$.

Suppose next that $A' = P^{-1}AP$ for some invertible matrix $n \times n$ matrix P. Then by N6.5.4 there exists a basis B' for V such that P is the change-of-basis matrix from B' to B. Since $A' = P^{-1}AP$ we conclude from 6.18 that A' is the matrix for A with respect to B'. □

N8.3 Diagonalization

Definition N8.3.1. *An $n \times n$-matrix A is called diagonal if $a_i = a_{ii}e_i$ for all $1 \leq i \leq n$.*

Lemma N8.3.2. *Let A be an $n \times n$-matrix. Then the following are equivalent*

(a) A is diagonal.

(b) $A = [d_1e_1, \dots, d_ne_n]$ for some $(d_1, \dots, d_n) \in \mathbb{R}^n$.

(c)

$$A = \begin{bmatrix} d_1 & 0 & 0 & \dots & 0 & 0 & 0 \\ 0 & d_2 & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & d_3 & \ddots & \ddots & 0 & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \ddots & \ddots & d_{n-2} & 0 & 0 \\ 0 & 0 & 0 & \ddots & 0 & d_{n-1} & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & d_n \end{bmatrix}$$

for some $(d_1, \dots, d_n) \in \mathbb{R}^n$

(d) $a_{ij} = 0$ for all $1 \leq i, j \leq n$ with $i \neq j$.

Proof. (a) $\Longrightarrow$ (b): Suppose A is diagonal and put $d_i = a_{ii}$. Then

$$A = [a_1, \dots, a_n] = [a_{11}e_1, \dots, a_{nn}e_n] = [d_1e_1, \dots, d_ne_n]$$

(b) $\Longrightarrow$ (c): Suppose (b) holds. Observe that

$$d_1e_1 = \begin{bmatrix} d_1 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad d_2e_2 = \begin{bmatrix} 0 \\ d_2 \\ 0 \\ \vdots \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \ldots \quad, \quad d_{n-1}e_{n-1} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ d_{n-1} \\ 0 \end{bmatrix}, \quad d_ne_n = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 0 \\ 0 \\ d_n \end{bmatrix}$$

and so (c) hold.

(c) $\Longrightarrow$ (d): Should be obvious.

(d) $\Longrightarrow$ (a): Suppose (d) holds and let $1 \leq i \leq n$ and $1 \leq j \leq n$. If $j \neq i$, then the j-entry of a_i is $a_{ij} = 0$ and the j-entry of $a_{ii}e_i$ is $a_{ii}0 = 0$. If $j = i$, then the j-entry of a_i is is a_{ii} and the j-entry of $a_{ii}e_i$ is $a_{ii}1 = a_{ii}$. So $a_i = a_{ii}e_i$ and A is diagonal. □

Definition N8.3.3. *(a) A square matrix is called diagonalizable if its is similar to a diagonal matrix.*

(b) Let $T : \mathbf{V} \to \mathbf{V}$ be linear and suppose that $\mathbf{V}$ is finite dimensional. Then T is called diagonalizable if there exists a basis B for $\mathbf{V}$ such that the matrix for T with respect to V is diagonal.

Lemma N8.3.4. *Let A be an $n \times n$-matrix. Suppose there exists a linearly independent list $B = (v_1, \ldots, v_n)$ in V such that for all $1 \leq i \leq n$, v_i is eigenvector of A associated to the eigenvalue λ_i of A. Put $P = [v_1, \ldots, v_n]$ and $D = [\lambda_1 e_1, \ldots, \lambda_n e_n]$. Then*

(a) B is a basis of $\mathbb{R}^n$.

(b) P is the change-of-basis matrix from B to the standard basis of $\mathbb{R}^n$.

(c) D is the matrix of L_A with respect to B.

(d) $D = P^{-1}AP$.

(e) A is diagonalizable.

Proof. (a) Since $\dim \mathbb{R}^n = n$ and B is a linearly independent list of length n in $\mathbb{R}^n$, N3.5.5 shows that B is basis.

(b) Let E be the standard basis for $\mathbb{R}^n$. Then by 6.15(b) the change-of-basis matrix from B to $\mathbb{R}^n$ is $\big[[v_1]_E, \ldots, [v_n]_E\big]$. Since $[x]_E = x$ for all $x \in \mathbb{R}^n$, we see that (b) holds.

(c) By definition, Column i of the matrix of L_A with respect to B is

$$\begin{aligned}
[L_A(v_i)]_B &= [Av_i]_B && \text{– definition of } L_A \\
&= [\lambda_i v_i]_B && \text{– since } v_i \text{ is an eigenvector of } A \text{ associated to } \lambda_i \\
&= \lambda_i [v_i]_B && \text{– } C_B \text{ is linear by N6.2.8} \\
&= \lambda_i e_i && \text{– N6.2.8} \\
&= d_i && \text{– definition of } D
\end{aligned}$$

and so (c) holds.

(d) By 6.10 A is the matrix of L_A with respect to E. So by (b) and 6.18 the matrix of L_A with respect to B is $P^{-1}AP$ and so (d) follows from (c).

(e) By (d), D is similar to A. Since $D = [\lambda_1 e_1, \ldots, \lambda_n e_n$, N8.3.2 shows that D is a diagonal matrix. Thus A is diagonalizable. □

Example N8.3.5. *Let* $A = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix}$. *Find a diagonal matrix* D *and an invertible matrix* P *with* $D = P^{-1}AP$.

$$\det \begin{bmatrix} \lambda & -1 & 0 \\ -1 & \lambda & 0 \\ 0 & 0 & \lambda+1 \end{bmatrix} = (\lambda+1)(\lambda^2 - (-1)(-1)) = (\lambda+1)(\lambda+1)(\lambda-1)$$

and so the eigenvalues are $\lambda = 1$ and $\lambda = -1$. We will use the Gauss Jordan Algorithm to compute a basis for $E_A(\lambda)$ for $\lambda = 1, -1$.

For $\lambda = 1$:

$$\begin{bmatrix} 1 & -1 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 2 \end{bmatrix} \xrightarrow[\substack{\frac{1}{2}R3 \to R3 \\ R2 \leftrightarrow R3}]{R2 + R1 \to R2} \begin{bmatrix} 1 & -1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

So x_2 is free, $x_1 = x_2$, $x_2 = x_2$ and $x_3 = 0$. Thus $\Big(1, 1, 0)\Big)$ is basis for $E_A(1)$.

For $\lambda = -1$:

$$\begin{bmatrix} -1 & -1 & 0 \\ -1 & -1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \xrightarrow[-R1 \to R1]{R2 - R1 \to R1} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

So x_2 and x_3 are is free, $x_1 = -x_2$, $x_2 = x_2$ and $x_3 = x_3$. Thus $\Big(-1,1,0),(0,0,1)\Big)$ is basis for $E_A(-1)$.

By 8.14,

$$B = \left(\begin{pmatrix} 1 \\ 1 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ -1 \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \right)$$

is linear independent. So we can apply N8.3.4. Put

$$P = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Then $D = P{-}1AP$. To verify this statement we will show that $PD = AP$:

$$PD = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

and

$$AP = \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Theorem 8.12. *Let $T : \mathbf{V} \to \mathbf{V}$ be linear, $B = (v_1, \ldots, v_n)$ a basis for V and A the matrix for T with respect to B. Let $\lambda \in \mathbb{R}$.*

(a) Let $v \in V$ and $x \in \mathbb{R}^n$ such that $x = [v]_B$ or $v = L_B(x)$. Then v is an eigenvector of T associated to λ if and only if x is an eigenvector for A associated to λ.

(b) The following three statements are equivalent:

(a) v_i is an eigenvector for T with respect to λ.

(b) $a_i = \lambda_i e_i$

(c) $a_i = a_{ii} e_i$ and $\lambda = a_{ii}$.

(c) A is diagonal if and only if for all $1 \leq i \leq n$, v_i is an eigenvector of T.

Proof. (a) Note first that by N6.2.8(a), $x = [v]_B$ if and only if $v = L_B(x)$.

$$\begin{array}{rrcll} & T(v)v & = & \lambda v & \\ \iff & [T(v)]_B & = & [\lambda v]_B & \text{– since } C_B \text{ is 1-1} \\ \iff & A[v]_B & = & \lambda[v]_B & \text{– 6.11(f), } C_B \text{ is linear} \\ \iff & Ax & = & \lambda x & \text{– since } x = [v]_B \end{array}$$

(b) Recall that $[v_i]_B = e_i$ by N6.2.8. Thus

$$\begin{array}{rcl} & v_i \text{ is an eigenvector for } T \text{ associated to } \lambda & \\ \iff & e_i \text{ is an eigenvector for } A \text{ associated to } \lambda & \text{– (a) and } [v_i]_B = e_i \\ \iff & Ae_i = \lambda e_i & \text{– definition of eigenvector} \\ \iff & a_i = \lambda e_i & \text{– since } Ae_i = a_i \text{ by N6.1.5} \end{array}$$

So (b:a) and (b:b) are equivalent.

If $a_i = \lambda e_i$, then a_{ii} is the i-entry of λe_i and so $a_{ii} = \lambda$. If $a_i = a_{ii}e_i$ and $\lambda = a_{ii}$, then $a_i \lambda e_i$. So (b:b) and (b:c) are equivalent.

(c) Follows from (b)

□

Theorem N8.3.7. *Let $T : \mathbf{V} \to \mathbf{V}$ be linear and suppose that $\mathbf{V}$ is finite dimensional. Then the following statements are equivalent:*

(a) T is diagonalizable.

(b) For each basis B of $\mathbf{V}$ the matrix for T with respect to B is diagonalizable.

(c) There exists a basis B for $\mathbf{V}$ such that the matrix for T with respect to B is diagonalizable.

(d) There exists a basis for $\mathbf{V}$ consisting of eigenvectors of T.

(e) The sum of the dimension of the eigenspaces of T equals the dimension of $\mathbf{V}$.

Proof. (a) $\Longrightarrow$ (b): Suppose T is diagonalizable, then there exists a basis F for $\mathbf{V}$ such that the matrix D for T with respect to F is diagonal. Let B be any basis for $\mathbf{V}$ and A the matrix for T with respect to B. By N8.2.2 B is similar to D and so (b) holds.

(b) $\Longrightarrow$ (c): Suppose that for each basis B of $\mathbf{V}$ the matrix for T with respect to B is diagonalizable. By N3.4.4 $\mathbf{V}$ has a basis B and so (c) holds.

(c) $\implies$ (a): Suppose there exists a B basis for $\mathbf{V}$ such that the matrix A for T with respect to B is diagonalizable. Then A is similar to a diagonal matrix D and by N8.2.2 there exists a basis B' of V such that the matrix of T with respect to B' is D. Thus T is diagonalizable.

(a) $\iff$ (d) : By definition T is diagonalizable if and only if there exists a basis B of V such that the matrix D of T with respect to B is diagonal. By 8.12 this holds if and only if there exists a basis for V consisting of eigenvectors of T. So (a) and (d) are equivalent

(d) $\iff$ (e) : Put $n = \dim \mathbf{V}$. Let $\lambda_1, \ldots, \lambda_k$ be the distinct eigenvalues for T and for $1 \leq i \leq k$ let $(u_{i1}, \ldots, u_{il_i})$ a basis of for $E_T(\lambda_i)$. By 8.14

$$B = (u_{11}, \ldots, u_{1l_1}, \ldots\ldots, u_{k1}, \ldots, u_{kl_k})$$

is linearly independent. Thus by the Comparison Theorem 3.9 $m := l_1 + l_2 + \ldots + l_k \leq n$.

(e) $\implies$ (d): Suppose $n = m$. Then by N3.5.5 B is basis for $\mathbf{V}$ and so (d) holds.

(d) $\implies$ (e): Let B be basis consisting of eigenvectors of $\mathbf{V}$ and for $1 \leq i \leq k$ let B_i be the sublist of B consisting of the elements of B associated to λ_i. Let n_i be the length of B_i. Then $n = n_1 + \ldots + n_k$. Note that B_i is a linearly independent list in $E_T(\lambda_i)$ and so $n_i \leq l_i$ by the Comparison Theorem 3.9. Thus

$$n = n_1 + \ldots + n_k \leq l_1 + \ldots + l_k = m \leq n$$

and $n = m$.

$\square$

Theorem 8.8. *Let A and A' be similar $n \times n$ matrices. Then*

(a) $\det A = \det A'$.

(b) For all $\lambda \in \mathbb{R}$, $\lambda I - A$ is similar to $\lambda I - A'$.

(c) A and A' have the same characteristic polynomial.

Proof. Let P be an invertible $n \times n$-matrix with $A' = P^{-1}AP$ and let $\lambda \in \mathbb{R}$.

(a) By Exercise 7.3.12, $\det A' = \det(P^{-1}AP) = \det A$.

(b) $P^{-1}(\lambda I - A)P = P^{-1}(\lambda I)P - P^{-1}AP = \lambda(P^{-1}IP) - A' = \lambda I - A'$.

(c) By (b), $\lambda I - A$ is similar to $\lambda I - A'$ and so by (a) $\det(\lambda I - A) = \det(\lambda I - A')$. Thus $\chi_A = \chi'_A$ by A.2.2. $\square$

Definition N8.3.9. *Let $T : \mathbf{V} \to \mathbf{V}$ be linear and suppose that V is finite dimensional. Then $\chi_T = \chi_A$ where A is the matrix of T with respect to some basis of V*

Note that this is well-defined by N8.2.2 and 8.8.

Appendix A

Functions and Relations

A.1 Basic definitions

Let n be a non-negative integer and a, b, c, d objects. Then (a, b) denotes the ordered pair formed by a and b. More formally, $(a, b) = \{\{a\}, \{a, b\}\}$, but we will never use this formal definition. Instead, we use the following fundamental property of ordered pairs:

$$(a, b) = (c, d) \text{ if and only if } a = c \text{ and } c = d$$

which can be proved from the definition and the axioms of Set Theory. a is called the first coordinate of (a, b) and b the second coordinate of (a, b).

(a, b, c) denotes the ordered triple formed by a, b and c. More formally $(a, b, c) = ((a, b), c)$.

Definition A.1.1. *Let A and B be sets.*

(a) $A \times B$ denotes the set

$$\{(a, b) \mid a \in A \text{ and } b \in B\}.$$

So the elements of $A \times B$ consists of all ordered pairs whose first coordinate is in A and the second is in B.

(b) A relation from A to B is a triple (A, B, R), denoted by $\sim$, such that R is a subset of $A \times B$. Let a and b be objects. We say that a is in $\sim$-relation to b and write $a \sim b$ if $(a, b) \in R$. So $a \sim b$ is a statement and

$$a \sim b \text{ if and only if } (a, b) \in R$$

(c) Let $\sim = (A, B, R)$ be a relation. A is called the domain of $\sim$ and B is called the codomain of $\sim$.

$$\operatorname{Im} \sim = \{b \in B \mid aRb \text{ for some } a \in A\},$$

$$\text{CoIm} \sim = \{a \in A \mid aRb \text{ for some } b \in B\}$$

Im $\sim$ *is called the Image of* $\sim$ *and* CoIm $\sim$ *the coimage of* $\sim$.

(d) A function from A to B is a relation F from A to B such that for all $a \in A$ there exists a unique b in B with aFb. We denote this unique b by Fa or by $F(a)$. So for $a \in A$ and $b \in B$,

$$b = Fa \text{ if and only if } aFb$$

Fa is called the image of a under F. If $b = Fa$ also will say that F maps a to b.

(e) We write "$F : A \to B$ is function" for "F is a function from A and B".

(f) Let $F : A \to B$ be a function and C a subset of A. Then $F[C] = \{F(c) \mid c \in C\}$.

Suppose for example that $A = \{1, 2, 3\}$ and $B = \{4, 5, 6\}$.

Put $R = \{(1, 4), (2, 5), (2, 6)\}$. Then $\sim = (A, B, R)$ is a relation from A to B with $1 \sim 4$, $2 \sim 5$ and $2 \sim 6$. But $\sim$ is not a function from A to B. Indeed, there does not exist an element b in R with $(1, b) \in R$. Also there exist two elements b in R with $(2, b) \in R$, namely $b = 5$ and $b = 6$.

Put $S = \{(1, 4), (2, 5), (3, 5)\}$. Then $F = (A, B, S)$ is the function from A to B with $F1 = 4$, $F2 = 5$ and $F3 = 5$.

Note that if $F = (A, B, R)$ is a function then Im $F = \{Fa \mid a \in A\}$ and CoIm $F = A$. Note that the text book uses the term range for the codomain of F. But since the term range is often used to denote the image of F, we prefer use the terms codomain and image.

Now let A and B be arbitrary sets and suppose that $\Phi(a)$ is a formula involving a variable a and if $a \in A$, then $\Phi(a)$ is in B. Put $R = \{(a, \Phi(a)) \mid a \in A\}$ and $F = (A, B, R)$. Then F is a function from A to B. We denote this function by

$$F : A \to B, a \to \Phi(a).$$

For example

$$F : \mathbb{R} \to \mathbb{R}, r \to r^2.$$

denotes the function from $\mathbb{R}$ to $\mathbb{R}$ with $Fr = r^2$ for all $r \in \mathbb{R}$.

A.2 Equality of relations

Lemma A.2.1. *Let A and B be sets.*

(a) Let $\sim = (A, B, R)$ be a relation from A to B. Then $R = \{(a, b) \mid a \in A, b \in B, a \sim b\}$.

(b) Let $\sim$ and $\approx$ be relations from A to B. Then $\sim=\approx$ if and only if for all $a \in A$ and $b \in B$ we have $a \sim b$ if and only if $a \approx b$.

Proof. (a) Put $S = \{(a,b) \mid a \in A, b \in B, a \sim b\}$. If $d \in S$, then by definition of S, $d = (a,b)$ for some $a \in A$, $b \in B$ and $a \sim b$. Hence by the definition of $a \sim b$, $(a,b) \in R$. Thus $d \in R$.

If $d \in R$ the since $R \subseteq A \times B$, $d = (a,b)$ for some $a \in A$ and $b \in B$. Since $(a,b) \in R$ we conclude that $a \sim b$ and so $d = (a,b) \in S$.

We proved that $d \in S$ if and only if $d \in R$ and so $R = S$.

(b) Let $\sim = (A,B,R)$ and $\approx = (A,B,T)$.

Suppose that $\approx=\sim$ and let $a \in A$ and $b \in B$. Then clearly $a \sim b$ if and only if $a \approx b$.

Suppose that for all $a \in A$ and $b \in B$ we have $a \sim b$ if and only if $a \approx b$. Then applying (a) to $\sim$ and $\approx$,

$$R = \{(a,b) \mid a \in A, b \in B, a \sim b\} = \{(a,b) \mid a \in A, b \in B, a \approx b\} = S$$

and so

$$\sim = (A,B,R) = (A,B,S) = \approx$$

□

Lemma A.2.2. *Let A and B sets and f and g functions from A to B. Then $f = g$ if and only if $fa = ga$ for all $a \in A$.*

Proof. If $f = g$, then clearly $fa = ga$ for all $a \in A$.

Suppose now that $fa = ga$ for all $a \in A$. Let $a \in A$ and $b \in B$.

$$\begin{array}{lcl} & afb & \\ \iff & b = fa & - \text{ definition of } fa \\ \iff & b = ga & - \text{ since } fa = ga \\ \iff & agb & - \text{ definition of } ga \end{array}$$

A.2.1 now show that $f = g$. □

A.3 Restriction of relations and function

This subsection has been used in earlier version of this lecture notes to treat subspaces, but currently is no longer used.

Lemma A.3.1. *Let $\sim$ be relation from A to B and C and D sets. Then then there exists a unique relation $\approx$ from C to D such that*

$$c \approx d \iff c \sim d$$

for all $c \in C$, $d \in D$.

Proof. We will first show that existence of $\approx$. Put $S = \{(c,d) \mid c \in C, d \in D, c \sim d\}$ and $\approx = (A,B,R)$. By definition of S, $S \subseteq C \times D$ and so $\approx$ is relation from C to D. Let $c \in C$ and $d \in D$.

Suppose that $c \approx d$. Then $(c,d) \in S$ and so by definition of S there exists $\tilde{c} \in C$ and $\tilde{d} \in D$ with $(c,d) = (\tilde{c},\tilde{d})$ and $\tilde{c} \sim \tilde{d}$. Hence $c = \tilde{c}, d = \tilde{d}$ and $c \sim d$.

Suppose that $c \sim d$. Then by definition of S, $(c,d) \in S$ and so $c \approx d$.

We proved that $c \approx d$ if and only if $c \sim d$ and the existence of $\approx$ is established.

Assume now that also $\simeq$ is a relation from C to B with

$$c \simeq d \Longleftrightarrow c \sim d$$

for all $c \in C$, $d \in D$.

Then for all $c \in C, d \in D$,

$$c \approx d \Longleftrightarrow c \sim d \Longleftrightarrow c \simeq d$$

and by A.2.1, $\approx = \simeq$. □

Definition A.3.2. *Let $\sim$ be relation from A to B, C and D sets and $\approx$ the unique relation from C to D such that $c \approx d \Longleftrightarrow c \sim d$ for all $c \in C$ and $d \in D$. Then $\approx$ is called the restriction of $\sim$ to C and D and is denoted by $\sim|_{C,D}$.*

Lemma A.3.3. *Let $f : A \to B$ be a function, C and D sets and $g = f\,|_{C,D}$ the restriction of f to C and D.*

(a) If g is a function, then $C \subseteq A$ and $gc = fc$ for all $c \in C$.

(b) g is a function if and only if $C \subseteq A$ and $fc \in D$ for all $c \in C$.

Proof. Suppose first that g is a function and let $c \in C$. Since g is a function, there exists a unique $d \in D$ with cgd. By definition of the restriction we conclude that cfd. In particular, $c \in A$ and so $C \subseteq A$. Moreover, by the definition of fc and gc we have $fc = d$ and $gc = d$. In particular, $fc = gc$ and $fc = d \in D$. So (a) is proved and also the forward direction of (b) is proved.

Suppose next that $C \subseteq A$ and $fc \in D$ for all $c \in C$. Let $c \in C$ and $d \in D$. Then by definition of g, cgd if and only if $c \in C, d \in D$ and cfd. Since $c \in C$ this is equivalent to $d \in D$ and cfd. Since f is a function this holds if and only if $d \in D$ and $d = fc$. Since $fc \in D$ for all $c \in C$ this holds if and only if $d = fc$. So there exists a unique element $d \in D$ with cgd (namely $d = fc$) and so g is a function. □

A.4 Composition of Relations

Definition A.4.1. *Let α be a relation from A to B and β a relation from B to C. Put*

$$S = \{(a,c) \in A \times C \mid (a\alpha b \text{ and } b\beta c) \text{ for some } b \in B\}$$

and

$$\beta \circ \alpha = (A, C, S).$$

Then $\beta \circ \alpha$ is called the composition of β and α.

Observe that $\beta \circ \alpha$ is a relation from A to C and if $a \in A$ and $c \in C$, then $a(\beta \circ \alpha)c$ if and only if there exists $b \in B$ with $a\alpha b$ and $b\beta c$.

Lemma A.4.2. *Let $f : A \to B$ and $g : B \to C$ be function. Then $g \circ f$ is a function and*

$$(g \circ f)a = g(fa)$$

for all $a \in A$.

Proof. Let $a \in A$, $b \in B$ and $c \in C$. Then

$$\begin{array}{rcl} & afb \text{ and } bgc & \\ \iff & b = fa \text{ and } c = gb & -\text{Definition of } fa, gb \\ \iff & b = fa \text{ and } c = g(fa) & -\text{Substitution} \end{array}$$

It follows that $a(g \circ f)c$ if and only if $c = g(fa)$. So $g \circ f$ is a function and $(g \circ f)a = g(fa)$. □

Lemma A.4.3. *Let $f : I \to J, g : J \to K$ and $h : K \to L$ be functions. Then $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. Let $i \in I$. Then

$$\begin{array}{rcl} & \Big(h \circ (g \circ f)\Big)i & \\ = & h\Big((g \circ f)i\Big) & -\text{ definition of composition} \\ = & h\Big(g(fi)\Big) & -\text{ definition of composition} \\ = & (h \circ g)(fi) & -\text{ definition of composition} \\ = & \Big((h \circ g) \circ f\Big)i & -\text{ definition of composition} \end{array}$$

Thus $h \circ (g \circ f) = (h \circ g) \circ f$ by A.2.2 □

Lemma A.4.4. *Let $f : I \to J, g : J \to K$ and $h : K \to L$ be relations. Then $h \circ (g \circ f) = (h \circ g) \circ f$.*

Proof. Let $i \in I$ and $l \in L$. Then

$$
\begin{array}{llll}
& i\Big(h \circ (g \circ f)\Big)l & & \\
\iff & \Big(i(g \circ f)k \text{ and } khl\Big) \text{ for some } k \in K & & - \text{ definition of composition} \\
\iff & \Big(\Big((ifj \text{ and } jgk) \text{ for some } j \in J\Big) \text{ and } khl\Big) \text{ for some } k \in K & & - \text{ definition of composition} \\
\iff & \Big(\Big((ifj \text{ and } jgk) \text{ and } khl\Big) \text{ for some } j \in J\Big) \text{ for some } k \in K & & - \text{(QR 10)} \\
\iff & \Big(\Big((ifj \text{ and } jgk) \text{ and } khl\Big) \text{ for some } k \in K\Big) \text{ for some } j \in J & & - \text{(QR 6)} \\
\iff & \Big(\Big(ifj \text{ and } (jgk \text{ and } khl)\Big) \text{ for some } k \in K\Big) \text{ for some } j \in J & & - \text{(LR 24)} \\
\iff & \Big(ifj \text{ and } \Big((jgk \text{ and } khl) \text{ for some } k \in K\Big)\Big) \text{ for some } j \in J & & - \text{(QR 10)} \\
\iff & \Big(ifj \text{ and } j(h \circ g)l\Big) \text{ for some } j \in J & & - \text{ definition of composition} \\
\iff & i\Big((h \circ g) \circ f\Big)l & & - \text{ definition of composition}
\end{array}
$$

Thus $h \circ (g \circ f) = (h \circ g) \circ f$ by A.2.1 □

A.5 Inverse of a function

Definition 6.4. *Let $f : A \to B$ be a function.*

(a) f is called 1-1 if, for all $b \in B$ there exists at most one $a \in A$ with $fa = b$. So f is 1-1 if and only

$$fa = fc \quad \Longrightarrow \quad a = c$$

for all $a, c \in A$.

(b) f is called onto if for all $b \in A$ there exists $a \in A$ with $b = fa$. So f is onto if and only if $B = \operatorname{Im} f$.

(c) An inverse of f is a function $g : B \to A$ such that

$$f \circ g = \mathrm{id}_B \text{ and } g \circ f = \mathrm{id}_B$$

(d) f is called invertible if there exists an inverse of f.

Lemma A.5.2. *Let $f : I \to J$ be a function. Then*

$$f \circ \mathrm{id}_I = f \text{ and } \mathrm{id}_J \circ f = f$$

Proof. See Homework 9 □

Lemma A.5.3. *Let $f : I \to J$ be an invertible function and f' an inverse of f.*

(a) Let $g : K \to I$ be a function, then $f' \circ (f \circ g) = g$.

(b) $g : K \to I$ and $h : K \to J$ be functions. Then

$$f \circ g = h \Longleftrightarrow g = f' \circ h$$

(c) There exists a unique function $f^* : J \to I$ with $f \circ f^* = \mathrm{id}_J$, namely $f^* = f'$.

(d) f' is the unique inverse of f.

(e) Let $g : J \to K$ be a function, then $(g \circ f) \circ f' = g$.

(f) Let $g : J \to K$ and $h : I \to K$ be functions. Then

$$g \circ f = h \Longleftrightarrow h = g \circ f'$$

(g) There exists a unique function $f^* : J \to I$ with $f^* \circ f = \mathrm{id}_I$, namely $f^* = f'$.

Proof. (a): $f' \circ (f \circ g) = (f' \circ f) \circ g = \mathrm{id}_I \circ g = g$.

(b): Suppose that $f \circ g = h$. Then using (a),

$$f' \circ h = f' \circ (f \circ g) = g$$

Suppose now that $g = f' \circ h$. Since f is an inverse of f' we can apply the result from the previous line and conclude that $h = f \circ g$. Thus (b) holds.

(c): Let $f^* : J \to I$ be a function. By (b) $f \circ f^* = \mathrm{id}_J$ if and only if $f^* = f$ $circ \mathrm{id}_J$, that is if and only if $f^* = f'$.

(d): This follows from (c).

(e): Similar to (a)

(f): Similar to (b).

(g): Similar to (c). □

Definition A.5.4. *Let $f : I \to J$ be an invertible function. Then f^{-1} denotes the unique inverse of f.*

Lemma A.5.5. *Let $f : A \to B$ and $g : B \to A$ be functions. Then the following four statements are equivalent:*

(a) g is an inverse of f.

(b) f is an inverse of g.

(c) $f(gb) = b$ for all $b \in B$ and $g(fa) = a$ for all $a \in A$.

(d) For all $a \in A$ and $b \in B$,

$$fa = b \quad \Longleftrightarrow \quad a = gb$$

Proof. (a) $\iff$ (b) : We have

$$
\begin{array}{lll}
 & g \text{ is an inverse of } f & \\
\iff & f \circ g = \mathrm{id}_B \text{ and } g \circ f = \mathrm{id}_A & - \text{ definition of inverse function} \\
\iff & g \circ f = \mathrm{id}_A \text{ and } f \circ g = \mathrm{id}_B & -(\text{LR } 13) \\
 & f \text{ is an inverse of } g & - \text{ definition of inverse function}
\end{array}
$$

So (a) and (b) are equivalent.

(b) $\iff$ (c) :

$$
\begin{array}{lll}
 & g \circ f = \mathrm{id}_A & \\
\iff & (g \circ f)a = \mathrm{id}_A a \text{ for all } a \in A & -\text{Equality of functions} \\
\iff & g(fa) = a \text{ for all } a \in A & -\text{Definition of composition and of } \mathrm{id}_A
\end{array}
$$

Similarly $f \circ g = \mathrm{id}_B$ if and only if $f(gb) = b$ for all $b \in B$. So (a) is equivalent to (c).

(c) $\implies$ (d): Suppose that (c) holds and let $a \in A$ and $b \in B$. If $fa = b$, then $a = g(fa) = gb$; and if $a = gb$, then $b = f(gb) = fa$. So $fa = b$ if and only if $a = gb$ and thus (d) holds

(d) $\implies$ (c): Suppose that (d) holds.

Let $a \in A$ and put $b = fa$. Then (d) implies $gb = a$ and so $g(fa) = a$.

Let $b \in B$ and put $a = gb$. Then (d) implies that $fa = b$ and so $f(gb) = b$. Thus (c) holds.

We proved that (c) implies (d) and that (d) implies (c). Hence (c) and (d) are equivalent.

□

Part (d) is a recipe for computing the inverse of a function $f : A \to B$. Consider for example the function $f : \mathbb{R} \to \mathbb{R}$ with $fx = 2x + 1$ for all $x \in \mathbb{R}$. Let $x, y \in \mathbb{R}$. Then

$$
\begin{array}{rccc}
 & fx & = & y \\
\iff & 2x + 1 & = & y \\
\iff & 2x & = & y - 1 \\
\iff & x & = & \frac{1}{2}(y - 1) \\
\iff & x & = & \frac{1}{2}y - \frac{1}{2}
\end{array}
$$

So the function $g : \mathbb{R} \to \mathbb{R}$ defined by $gy = \frac{1}{2}y - \frac{1}{2}$ for all $y \in \mathbb{R}$ is an inverse for f.

Lemma A.5.6. *Let $f : A \to B$ and $g : B \to C$ be invertible. Then $g \circ f$ is invertible with inverse $f^{-1} \circ g^{-1}$.*

Proof. Let $a \in A, c \in C$. Then

$$\begin{array}{rrcll} & (g \circ f)a & = & c & \\ \iff & g(fa) & = & c & -\text{ definition of composition} \\ \iff & fa & = & g^{-1}c & -\text{ A.5.5 applied to the inverse of } g \\ \iff & a & = & f^{-1}(g^{-1}c) & -\text{ A.5.5 applied to the inverse of } f \\ \iff & a & = & (f^{-1} \circ g^{-1})c & -\text{ definition of composition} \end{array}$$

So by A.5.5 $f^{-1} \circ g^{-1}$ is the inverse of $g \circ f$. □

Theorem 6.6. *Let f be a function. Then f is invertible if and only if f is 1-1 and onto. That is $f : A \to B$ is invertible if and only if for all $b \in B$ there exists a unique $a \in A$ with $fa = b$.*

Proof. $\Longrightarrow$: Suppose first that f is invertible and let g be an inverse of f. Let $a, c \in A$.

$$\begin{array}{rcl} & fa = fc & \\ \Longrightarrow & g(fa) = g(fc) & -\text{Substitution} \\ \Longrightarrow & a = c & -\text{A.5.5, twice} \end{array}$$

Thus f is 1-1. Now let $b \in B$ and put $a = gb$. Then $a \in A$ and by A.5.5(c), $fa = b$ and f is onto.

$\Longleftarrow$: I will give two proofs for the backward direction:

Proof 1: Suppose that f is 1-1 and onto. Since f is onto, we can choose for each $b \in B$ an element $b' \in A$ with $fb' = b$. Define $g : B \to A$ by $gb = b'$. Let $a \in A$ and $b \in B$. Then

$$\begin{array}{rcl} & fa = b & \\ \iff & fa = fb' & -\text{ since } fb' = b \\ \iff & a = b' & -\ f \text{ is 1-1} \\ \iff & a = gb & -\text{ definition of } g \end{array}$$

So by A.5.5(c), g is an inverse of f.

Proof 2: Suppose that f is 1-1 and onto. Put $S = \{(fa, a) \mid a \in A\}$. Then $S \subseteq B \times A$ and so $g := (B, A, S)$ is a relation. Let $a \in A$ and $b \in B$. Then

$$\begin{array}{lrcl} & & bga & \\ (*) & \iff & (b, a) \in S & -\text{ Definition of } bga \\ & \iff & b = fa & -\text{ Definition of } S \end{array}$$

Let $b \in B$. Since f is onto, there exists $a \in A$ with $b = fa$ and so by (*), bga. Let $a, c \in A$ with bga and bgc. Then by (*), $fa = b = fc$ and since f is 1-1, $a = c$. So for each $b \in B$ there exists a unique $a \in A$ with bga. Hence g is a function.

Let $a \in A$ and $b \in B$. Since g is a function, $a = gb$ if and only if bga So (*) implies that

$$a = gb \quad \Longleftrightarrow \quad b = fa$$

Thus by A.5.5, g is an inverse of f. □

Lemma A.5.8. *Let $f : A \to B$ be an invertible function with inverse $g : B \to A$. Let $C \subseteq A$ and $D \subseteq A$. Then the following are equivalent:*

(a) $f[C] \subseteq D$ and $g[D] \subseteq C$.

(b) $f\,|_{C,D}$ and $g\,|_{D,C}$ are functions.

(c) $f\,|_{C,D}$ and $g\,|_{C,D}$ are functions inverse to each other.

(d) $f\,|_{C,D}$ is an invertible function.

(e) $f[C] = D$.

(f) For all $a \in A$, $a \in C$ if and only if $f(a) \in D$.

(g) $g\,|_{D,C}$ is an invertible function.

(h) $g[D] = C$

(i) For all $b \in B$, $g(b) \in C$ if and only if $b \in B$.

Proof. Put $\tilde{f} = f\,|_{C,D}$ and $\tilde{g} = g\,|_{D,C}$

(a) $\Longrightarrow$ (b): This follows from A.3.3.

(b) $\Longrightarrow$ (c): By A.3.3 we get $\tilde{f}(\tilde{g}d) = f(gd) = d$ for all $d \in D$ and $\tilde{g}(\tilde{f}c) = g(fc)$ for all $c \in C$. Thus $\tilde{f}$ is the inverse of $\tilde{g}$.

(c) $\Longrightarrow$ (d): This implication follows from the definition of invertible.

(d) $\Longrightarrow$ (e): Since $\tilde{f}$ is a function, A.3.3 gives $fc = \tilde{f}c$ for all $c \in C$. Hence

$$f[C] = \{fc \mid c \in C\} = \{\tilde{f}c \mid c \in C\} = \tilde{f}[C].$$

Since $\tilde{f}$ is invertible, 6.6 shows that $\tilde{f}$ is onto. So $\tilde{f}[C] = D$ and then $f[C] = D$.

(e) $\Longrightarrow$ (a): Suppose $f[C] = D$. Then clearly $f[C] \subseteq D$. Let $d \in D$. Since $f[C] = D$, $d = fc$ for some $c \in C$ and so $gd = c \in C$. Thus also $f[C] \subseteq D$.

Thus the first five statements are equivalent.

(a) $\Longrightarrow$ (f): Suppose (a) holds. Let $a \in A$. If $a \in C$ then $fa \in f[C] \subseteq D$. And if $fa \in D$, then $a = g(fa) \in g[D] \subseteq C$.

(f) $\Longrightarrow$ (a): Suppose (f) holds. Then clearly $f[C] \subseteq D$. Let $d \in D$. Then $f(gd) = d \in D$ and so $gd \in C$ since (f) holds. Thus $g[D] \subseteq C$ and (a) is proved.

So also (f) is equivalent to (a).

We proved that (d), (e) and (f) are equivalent to (a). This result applied with the roles of f and g interchanged shows that also (g), (h) and (i) are equivalent of (a). □

A.6 Defining Sequences by Induction

Theorem A.6.1. *Let I be a non-empty set, $f : I \to I$ a function and $d \in I$. Then there exists a unique sequence $a = \big(a(n)\big)_{n=1}^{\infty}$ such that*

(i) $a(1) = d$, *and*

(ii) $a(n+1) = f\big(a(n)\big)$ *for all* $n \in \mathbb{N}$.

Proof. For $m \in \mathbb{N}$ let S_m be the following statement:

There exists a unique list $e_m = \big(e_m(n)\big)_{n=1}^{m}$ *of length m in I such that*

(i') $e_m(1) = d$.

(ii') $e_m(n+1) = f\big(e_m(n)\big)$ *for all* $1 \leq n < m$.

Note that S_1 holds with $e_1 = (d)$.

Suppose now that S_k holds. So there exists a unique list e_k which fulfils (i') and (ii') for $m = k$. Define the list e_{k+1} of length $k+1$ in I by

$$(*) \qquad e_{k+1}(n) = \begin{cases} e_k(n) & \text{if } 1 \leq n \leq k \\ f\big(e_k(k)\big) & \text{if } n = k+1 \end{cases}$$

Let $e = \big(e(n)\big)_{n=1}^{k+1}$ be a list of length $k = 1$ in I. Observe that e fulfill (i') and (ii') for $m = k+1$ if and only if

(i") $e(1) = d$.

(ii") $e(n+1) = f\big(e(n)\big)$ for all $1 \leq n < k$.

(iii") $e(k+1) = f\big(e(k)\big)$

By the induction assumption (i") and (ii") hold if and only if $e(n) = e_k(n)$ for all $1 \leq n \leq k$ and so (i")- (iii") hold if and only if in addition $e(k+1) = f\big(e_k(k)\big)$. So e_{k+1} is the unique list of length $k+1$ which fulfills (i') and (ii').

Thus S_{k+1} holds and by the Principal of Mathematical Induction we conclude that S_m holds for all $m \in \mathbb{N}$.

Now let $b = \big(b(n)\big)_{n=1}^{\infty}$ be sequence which fulfills (i) and (ii). Observe that $b_m = \big(b(n)\big)_{n=1}^{m}$ fulfills (i') and (ii'). So the uniqueness assertions in S_m implies $b_m = e_m$. In particular

$$b(m) = e_m(m) \text{ for all } 1 \leq m < \infty$$

Thus b is uniquely determined. Conversely define the infinite list $a = \big(a(n)\big)_{n=1}^{\infty}$ via

$$a(n) = e_n(n) \text{ for all } 1 \leq n < \infty$$

Then $a(1) = e_1(1) = d$ and using (*) for $k = n$

$$a(n+1) = e_{n+1}(n+1) = f\big(e_n(n)\big) = f\big(a(n)\big).$$

So (i) and (ii) holds for a and so a is the unique sequence which fulfills (i) and (ii). $\square$

Appendix B

Logic

B.1 Quantifiers

Let P be a statement involving a variable x. Then

$$\forall x(P) \text{ is the statement that } P \text{ is true for all objects } x.$$

Note hat $\forall x(P)$ is false if there exists an object x such that P is false. Applying this to $\neg P$ instead of P we see that $\forall x(\neg P)$ is false if there exists an object x such that P is true. We use this observation to define the statement $\exists x(P)$ to be $\neg(\forall x(\neg P))$. So

$$\exists x(P) \text{ is the statement that there exists an object } x \text{ such that } P \text{ is true.}$$

The symbols $\forall$ and $\exists$ are called quantifiers. The following theorems list a few statements involving quantifiers which are always true.

Theorem B.1.1. *Let P and Q be statements and x and y variables.*

QR 1 $\neg\Big(\forall x(\neg P)\Big) \iff \exists x(P)$.

QR 2 $\neg\Big(\forall x(P)\Big) \iff \exists x(\neg P)$.

QR 3 $\forall x(P) \iff \neg\Big(\exists x(\neg P)\Big)$.

QR 4 $\forall x(\neg P) \iff \neg\Big(\exists x(P)\Big)$.

QR 5 $\forall x\Big(\forall y(P)\Big) \iff \forall y\Big(\forall x(P)\Big)$.

QR 6 $\exists x\Big(\exists y(P)\Big) \iff \exists y\Big(\exists x(P)\Big)$.

QR 7 $\forall x(P \text{ and } Q) \iff \Big(\big(\forall x(P)\big) \text{ and } \big(\forall x(Q)\big)\Big)$

QR 8 *If Q does not involve x, then*

$$\forall x(P \text{ or } Q) \iff \Big(\big(\forall x(P)\big) \text{ or } Q\Big).$$

QR 9 $\exists x(P \text{ or } Q) \iff \Big(\big(\exists x(P)\big) \text{ or } \big(\exists x(Q)\big)\Big)$

QR 10 *If Q does not involve x, then*

$$\exists x(P \text{ and } Q) \iff \Big(\big(\exists x(P)\big) \text{ and } Q\Big)$$

The statement $\forall(x \in I)(P)$ is defined as $\forall x\big((x \in I) \implies P\big)$. The statement $\exists(x \in I)(P)$ is defined as $\exists x\big((x \in I) \text{ and } P\big)$. Note that

$$\neg\Big(\exists(x \in I)(P)\Big) \iff \Big(\forall(x \in I)(\neg P)\Big)$$

Indeed

$$\begin{array}{cc} & \neg\Big(\exists(x \in I)(P)\Big) \\ \iff & \neg\Big(\exists x\big((x \in I) \text{ and } P\big)\Big) \\ \iff & \forall x\Big(\neg\big((x \in I) \text{ and } P\big)\Big) \\ \iff & \forall x\Big(\neg(x \in I) \text{ or } \neg P\Big) \\ \iff & \forall x\Big((x \in I) \implies \neg P\Big) \\ \iff & \forall(x \in I)(\neg P) \end{array}$$

Then writing proofs we will rarely use the symbols $\forall$ and $\exists$, but rather use phrases like "for all x", "there exists x" or "for some x".

Appendix C

The real numbers

C.1 Definition

Definition C.1.1. *The real numbers are a quadtruple* $(\mathbb{R}, +, \cdot, \leq)$ *such that*

($\mathbb{R}$ i) $\mathbb{R}$ *is a set (whose elements are called* real numbers*)*

($\mathbb{R}$ ii) $+$ *is a function (called* addition*) ,* $\mathbb{R} \times \mathbb{R}$ *is a subset of the domain of* $+$ *and*

$$a + b \in \mathbb{R} \qquad \text{(Closure of addition)}$$

for all $a, b \in \mathbb{R}$*, where* $a \oplus b$ *denotes the image of* (a, b) *under* $+$*;*

($\mathbb{R}$ iii) $\cdot$ *is a function (*called *multiplication),* $\mathbb{R} \times \mathbb{R}$ *is a subset of the domain of* $\cdot$ *and*

$$a \cdot b \in \mathbb{R} \qquad \text{(Closure of multiplication)}$$

for all $a, b \in \mathbb{R}$ *where* $a \cdot b$ *denotes the image of* (a, b) *under* $\cdot$*. We will also use the notion* ab *for* $a \cdot b$*.*

($\mathbb{R}$ iv) $\leq$ *is a relation between* $\mathbb{R}$ *and* $\mathbb{R}$*;*

and such that the following statements hold:

($\mathbb{R}$ Ax 1) $a + b = b + a$ *for all* $a, b \in \mathbb{R}$. (Commutativity of Addition)

($\mathbb{R}$ Ax 2) $a + (b + c) = (a + b) + c$ *for all* $a, b, c \in \mathbb{R}$*;* (Associativity of Addition)

($\mathbb{R}$ Ax 3) *There exists an element in* $\mathbb{R}$*, denoted by* 0 *(and called* zero*), such that* $a + 0 = a$ *and* $0 + a = a$ *for all* $a \in \mathbb{R}$*;* (Existence of Additive Identity)

($\mathbb{R}$ Ax 4) *For each* $a \in \mathbb{R}$ *there exists an element in* $\mathbb{R}$*, denoted by* $-a$ *(and called negative a) such that* $a + (-a) = 0$ *and* $(-a) + a = 0$*;* (Existence of Additive Inverse)

($\mathbb{R}$ Ax 5) *$a(b+c) = ab + ac$ for all $a, b, c \in \mathbb{R}$.* (Right Distributivity)

($\mathbb{R}$ Ax 6) *$(a+b)c = ac + bc$ for all $a, b, c \in \mathbb{R}$* (Left Distributivity)

($\mathbb{R}$ Ax 7) *$(ab)c = a(bc)$ for all $a, b, c \in \mathbb{R}$* (Associativity of Multiplication)

($\mathbb{R}$ Ax 8) *There exists an element in $\mathbb{R}$, denoted by 1 (and called* one*), such that $1a = a$ for all $a \in R$.* (Multiplicative Identity)

($\mathbb{R}$ Ax 9) *For each $a \in \mathbb{R}$ with $a \neq 0$ there exists an element in $\mathbb{R}$, denoted by $\frac{1}{a}$ (and called 'a* inverse*') such that $aa^{-1} = 1$ and $a^{-1}a = 1$;*

(Existence of Multiplicative Inverse)

($\mathbb{R}$ Ax 10) *For all $a, b \in \mathbb{R}$,*

$$(a \leq b \text{ and } b \leq a) \Longleftrightarrow (a = b)$$

($\mathbb{R}$ Ax 11) *For all $a, b, c \in \mathbb{R}$,*

$$(a \leq b \text{ and } b \leq c) \Longrightarrow (a \leq c)$$

($\mathbb{R}$ Ax 12) *For all $a, b, c \in \mathbb{R}$,*

$$(a \leq b \text{ and } 0 \leq c) \Longrightarrow (ac \leq bc)$$

($\mathbb{R}$ Ax 13) *For all $a, b, c \in \mathbb{R}$,*

$$(a \leq b) \Longrightarrow (a + c \leq b + c)$$

($\mathbb{R}$ Ax 14) *Each bounded, non-empty subset of $\mathbb{R}$ has a least upper bound. That is, if S is a non-empty subset of $\mathbb{R}$ and there exists $u \in \mathbb{R}$ with $s \leq u$ for all $s \in S$, then there exists $m \in R$ such that for all $r \in \mathbb{R}$,*

$$\Big(s \leq r \text{ for all } s \in S\Big) \Longleftrightarrow \Big(m \leq r\Big)$$

($\mathbb{R}$ Ax 15) *For all $a, b \in \mathbb{R}$ such that $b \neq 0$ and $0 \leq b$ there exists a positive integer n such that $a \leq nb$. (Here na is inductively defined by $1a = a$ and $(n+1)a = na + a$).*

Definition C.1.2. *The relations $<$, $\geq$ and $>$ on $\mathbb{R}$ are defined as follows: Let $a, b \in \mathbb{R}$, then*

(a) $a < b$ if $a \leq b$ and $a \neq b$.

(b) $a \geq b$ if $b \leq a$.

(c) $a > b$ if $b \leq a$ and $a \neq b$

Appendix D

General Commutative and Associative Laws

D.1 Sums

Lemma D.1.1. *Let* $\mathbf{V}$ *be a vector space.*

(a) Let $(v_1, \ldots v_n)$ *and* $(w_1, \ldots, w_m)$ *be list in* V*. Then*

$$(v_1 + \ldots + v_n) + (w_1 + \ldots + w_m) = v_1 + \ldots + v_n + w_1 + \ldots + w_m$$

(b) Let $(v_1, \ldots v_n)$ *and* $(w_1, \ldots, w_n)$ *be lists of the same length in* V*. Then*

$$(v_1 + \ldots + v_n) + (w_1 + \ldots + w_n) = (v_1 + w_1) + \ldots + (v_n + w_n)$$

(c) Let $(v_1, \ldots v_n)$ *be a list in* V *and* $r \in \mathbb{R}$*. Then*

$$r(v_1 + \ldots + v_n) = rv_1 + \ldots + rv_n$$

Proof. (a) The proof is by induction on m. For $m = 0$ the left side in (a) is $(v_1 + \ldots + v_n) + \mathbf{0}$ and the right side is $v_1 + \ldots + v_n$. So by (Ax 3) (a) holds for $m = 0$. Suppose now that (a) holds for m. Then

$$\begin{array}{rll}
& (v_1 + \ldots + v_n) + (w_1 + \ldots + w_{m+1})) & \\
= & \Big(v_1 + \ldots + v_n) + v_{n+1}\Big) + \Big((w_1 + \ldots + w_n) + w_{n+1}\Big) & - \text{ definition of } '+ \ldots +' \\
= & \Big((v_1 + \ldots + v_n) + (w_1 + \ldots + w_n)\Big) + w_{n+1} & - \text{ (Ax 2)} \\
= & \Big(v_1 + \ldots + v_n + w_1 + \ldots + w_m\Big) + w_{n+1} & - \text{ Induction assumption} \\
= & v_1 + \ldots + v_n + w_1 + \ldots + w_{m+1} & - \text{ definition of } '+ \ldots +.'
\end{array}$$

So the (a) holds for $m+1$ and thus by the principal of induction for all non-negative integers m.

(b) The proof is by induction on n. For $n=0$ the left side in (b) is $\mathbf{0}+\mathbf{0}$ and the right side is $\mathbf{0}$ and so by (Ax 3) (b) holds for $n=0$. Suppose now that the lemma holds for n. Then

$$\begin{aligned}
& (v_1+\ldots+v_{n+1})+(w_1+\ldots+w_{n+1}) \\
= \ & \Big((v_1+\ldots+v_n)+v_{n+1}\Big)+\Big(w_1+\ldots+w_{n+1}\Big) && -\text{ definition of }'+\ldots+' \\
= \ & \Big((v_1+\ldots+v_n)+v_{n+1}\Big)+\Big((w_1+\ldots+w_n)+w_{n+1}\Big) && -\text{ definition of }'+\ldots+.' \\
= \ & \Bigg(\Big(v_1+\ldots+v_n)+v_{n+1}\Big)+\Big(w_1+\ldots+w_n\Big)\Bigg)+w_{n+1} && -\text{ (Ax 2)} \\
= \ & \Bigg((v_1+\ldots+v_n)+\Big(v_{n+1}+(w_1+\ldots+w_n)\Big)\Bigg)+w_{n+1} && -\text{ (Ax 2)} \\
= \ & \Bigg((v_1+\ldots+v_n)+\Big((w_1+\ldots+w_n)+v_{n+1}\Big)\Bigg)+w_{n+1} && -\text{ (Ax 1)} \\
= \ & \Bigg(\Big((v_1+\ldots+v_n)+(w_1+\ldots+w_n)\Big)+v_{n+1}\Bigg)+w_{n+1} && -\text{ (Ax 2)} \\
= \ & \Big((v_1+\ldots+v_n)+(w_1+\ldots+w_n)\Big)+\Big(v_{n+1}+w_{n+1}\Big) && -\text{ (Ax 2)} \\
= \ & \Big((v_1+w_1)+\ldots+(v_n+w_n)\Big)+\Big(v_{n+1}+w_{n+1}\Big) && -\text{ Induction assumption} \\
= \ & (v_1+w_1)+\ldots+(v_{n+1}+w_{n+1}) && -\text{ definition of }'+\ldots+.'
\end{aligned}$$

Hence (b) holds for $n+1$ and thus by the principal of induction for all non-negative integers n.

(c) The proof is by induction on n. For $n=0$ the left side in (c) is $r\mathbf{0}$ and the right side is $\mathbf{0}$. So by 1.4 (c) holds for $n=0$. Suppose now that (c) holds for n. Then

$$\begin{aligned}
& r(v_1+\ldots+v_{n+1}) \\
= \ & r\Big((v_1+\ldots+v_n)+v_{n+1}\Big) && -\text{ definition of }'+\ldots+' \\
= \ & r(v_1+\ldots+v_n)+rv_{n+1} && -\text{ (Ax 5)} \\
= \ & (rv_1+\ldots+rv_n)+rv_{n+1} && -\text{ Induction assumption} \\
= \ & rv_1+\ldots+rv_{n+1} && -\text{ definition of }'+\ldots+.'
\end{aligned}$$

Hence (c) holds for $n+1$ and thus by the principal of induction for all non-negative integers n $\square$

D.2 Linear combinations

Lemma D.2.1. *Let* $\mathbf{V}$ *be a vector space,* $(v_1, \dots v_n)$ *a list in* V *and* $(r_1, \dots, r_n)$ *a list in* $\mathbb{R}^n$.

(a) Let $(s_1, \dots, s_n)$ *be list in* $\mathbb{R}$. *Then*

$$(r_1 v_1 + \dots + r_n v_n) + (s_1 v_1 + \dots + s_n v_n) = (s_1 + r_1)v_1 + \dots + (s_n + r_n)v_n$$

(b) Let $s \in \mathbb{R}$. *Then*

$$s(r_1 v_1 + \dots + r_n v_n) = (sr_1)v_1 + \dots + (sr_n)v_n$$

Proof. (a): By D.1.1(b) $(r_1 v_1 + \dots + r_n v_n) + (s_1 v_1 + \dots + s_n v_n) = (r_1 v_1 + s_1 v_1) + \dots + (r_n v_n + s_n v_n)$. By (Ax 6) the latter is equal to $(s_1 + r_1)v_1 + \dots + (s_n + r_n)v_n$.

(b) By D.1.1(c) $s(r_1 v_1 + \dots + r_n v_n) = s(r_1 v_1) + \dots + s(r_n v_n)$. By (Ax 7) the latter is equal to $(sr_1)v_1 + \dots + (sr_n)v_n$. $\square$

Bibliography

[1] Robert Messer *Linear Algebra, Gateway to Mathematics*, Harper Collins College Publisher **1993**.

Index