DON'T PANIC: when in doubt, row-reduce!

A hitch hiker's guide to linear algebra

Spring Term 2008

Contents

1 Greetings

The goal of this class is to make you understand the joke in the title.

2 Preliminaries on Algebra

Definition 2.1. Let M be a set. A law of composition $\otimes$ on M is a map

$$\begin{aligned} \otimes : M \times M &\longrightarrow M \\ (m,n) &\mapsto m \otimes n \end{aligned}$$

This is the mathematicians way to say that a law of composition takes in two elements of M and returns (depending on its input) another element of M. A law of composition may be sensitive to the order of its arguments.

Example 2.2. Taking differences is a law of composition on the set of real numbers:

$$\begin{aligned} - : \mathbb{R} \times \mathbb{R} &\longrightarrow \mathbb{R} \\ (t,s) &\mapsto t - s \end{aligned}$$

Note that $5 - 3 \neq 3 - 5$.

All algebra starts with a set that carries a law of composition. Such a mathematical structure is called a magma, which tells you something about mathematical humor. More interesting structures are obtained by imposing further restrictions. (Almost nothing can be said about magmas.)

2.1 Groups

Groups are the most interesting things in mathematics (I am a group theorist, just in case you wonder). They arise from magmas via three innocent looking requirements:

Definition 2.3. A group is a set G together with a distinguished element (called identity-element) I and a law of composition

$$\begin{aligned} \otimes : G \times G &\rightarrow G \\ (g,h) &\mapsto g \otimes h \end{aligned}$$

such that the following three axioms hold:

1. The operation $\otimes$ is associative, i.e., we have

   $$(g_1 \otimes g_2) \otimes g_3 = g_1 \otimes (g_2 \otimes g_3) \qquad \text{for all } g_1, g_2, g_3 \in G.$$

2. The element I is a right-identity element, i.e., we have

   $$g \otimes I = g \qquad \text{for all } g \in G.$$

3. Every group element g has a right-inverse, i.e., there is an element j satisfying

   $$g \otimes j = I.$$

   (Of course different elements may have different inverses.)

Theorem 2.4. *Let $(G, I, \otimes)$ be a group.*

1. *Right-inverses are left-inverses, i.e., we have*

$$\text{if } g \otimes j = I \text{ then } j \otimes g = I$$

   for any two $g, j \in G$.

2. *The right-identity I is a left-identity, i.e., we have*

$$I \otimes g = g$$

   for every $g \in G$.

Proof. (1) We assume $g \otimes j = I$. Now, we calculate:

$$\begin{array}{lll} j \otimes g & = (j \otimes g) \otimes I & \text{by axiom 2} \\ & = (j \otimes g) \otimes (j \otimes k) & \text{by axiom 3, where } k \text{ is a right-inverse for } j \\ & = ((j \otimes g) \otimes j) \otimes k & \text{by axiom 1} \\ & = (j \otimes (g \otimes j)) \otimes k & \text{by axiom 1} \\ & = (j \otimes I) \otimes k & \text{by our hypothesis that } g \otimes j = I \\ & = j \otimes k & \text{by axiom 2} \\ & = I & \text{recall that } k \text{ is a right-inverse for } j \end{array}$$

(2) Let g be any element of G. We pick an inverse j, and by part 1, which has been proved already, we know that

$$g \otimes j = I = j \otimes g.$$

With this information, we can carry out the calculation:

$$\begin{array}{lll} I \otimes g & = (g \otimes j) \otimes g & \text{since } g \otimes j = I \\ & = g \otimes (j \otimes g) & \text{by axiom 1} \\ & = g \otimes I & \text{since } j \otimes g = I \text{ by part 1} \\ & = g & \text{by axiom 2} \end{array}$$

q.e.d.

Theorem 2.5. *Let $(G, I, \otimes)$ be a group, and fix two elements $g, h \in G$. Then:*

1. *Left-quotients exist and are unique, i.e. the equation*

$$x \otimes g = h$$

   has a solution, and this solution is unique.

2. *Right-quotients exist and are unique, i.e. the equation*

$$g \otimes y = h$$

   has a solution, and this solution is unique.

Warning: *It is not necessarily true that the left- and the right-quotient coincide.*

Proof. We only proof (1). Since Theorem 2.4 establishes a symmetry of left and right, we can leave the proof for (2) as an exercise.

We deal with uniqueness first. So let us assume that

$$\begin{aligned} x_0 \otimes g &= h \\ x_1 \otimes g &= h \end{aligned}$$

Uniqueness of solutions means that we have to show $x_0 = x_1$. Let j be a (right-) inverse for g, i.e., pick j so that $g \otimes j = I$. Then we have:

$$\begin{array}{rl|l} x_0 \otimes g & = h & \text{by hypothesis} \\ (x_0 \otimes g) \otimes j & = h \otimes j & \text{by applying } \otimes j \\ x_0 \otimes (g \otimes j) & = h \otimes j & \text{by axiom 1} \\ x_0 \otimes I & = h \otimes j & \text{by axiom 3} \\ x_0 & = h \otimes j & \text{by axiom 2} \end{array}$$

By the very same reasoning, we obtain $x_1 = h \otimes j$. Hence

$$x_0 = h \otimes j = x_1,$$

which settles uniqueness.

To prove the existence of a solution, we check the candidate that we obtained in our uniqueness proof: We claim that $x = h \otimes j$ will solve the equation

$$x \otimes g = h$$

We verify this:

$$\begin{array}{rl|l} x \otimes g & = (h \otimes j) \otimes g & \text{by our choice } x = h \otimes j \\ & = h \otimes (j \otimes g) & \text{by axiom 1} \\ & = h \otimes I & \text{by Theorem 2.4, } j \text{ is a left-inverse for } g \\ & = h & \text{by axiom 2} \end{array}$$

Note how it becomes important that the right-inverse j for g, which we picked in the uniqueness argument, is a left-inverse when we want to prove that the only possible candidate for a solution actually is a solution. **q.e.d.**

Corollary 2.6. *Inverses are unique: The right-inverse for g solves the equation $g \otimes x = I$. Since there is only one such x, right-inverses are unique. We denote the right-inverse for g by $\overline{g}$.*

Corollary 2.7. *The identity element is unique, i.e., if an element satisfies $g \otimes x = g$ then $x = I$. The reason is that I satisfies the equation, and any equation of that sort has but one solution.*

Exercise 2.8. Let $(G, I, \otimes)$ be a group, and let g and h be elements of G. Show that

$$\overline{g \otimes h} = \overline{h} \otimes \overline{g}.$$

Exercise 2.9. Addition of integers induces a well-defined way of "adding parities":

+	even	odd
even	even	odd
odd	odd	even

Put $G := \{\text{even}, \text{odd}\}$. Show that $(G, \text{even}, +)$ is a group.

Exercise 2.10. Let G be the set of non-zero real numbers. We define a law of composition on G as follows:

$$g \otimes h := |g|h.$$

Show that

1. The non-zero real number $1 \in G$ is a left-identity element for $\otimes$ in G.
2. Every element $g \in G$ has a right-inverse with respect to $\otimes$, i.e., there is an $j \in G$ such that
$$g \otimes j = 1.$$
3. The structure $(G, 1, \otimes)$ is <u>not</u> a group.

2.2 Fields

Definition 2.11. A field is a set $\mathbb{F}$ together with two laws of composition $+$ (called addition) and $\cdot$ (called multiplication) and two distinguished elements 0 (called zero) and 1 (called one), such that the following axioms hold:

1. Addition in $\mathbb{F}$ obeys the law of associativity:
$$(r+s)+t = r+(s+t) \qquad \text{for all } r,s,t \in \mathbb{F}.$$

2. The zero-element is a right-identity for addition:
$$r+0 = r \qquad \text{for all } r \in \mathbb{F}.$$

3. Addition in $\mathbb{F}$ admits right-inverses: for every $r \in \mathbb{F}$, there is an element $s \in \mathbb{F}$ such that
$$r+s = 0.$$

4. Multiplication in $\mathbb{F}$ obeys the law of associativity:
$$(r \cdot s) \cdot t = r \cdot (s \cdot t) \qquad \text{for all } r,s,t \in \mathbb{F}.$$

5. The one-element is a right-identity for multiplication:
$$r \cdot 1 = r \qquad \text{for all } r \in \mathbb{F}.$$

6. Non-zero elements in $\mathbb{F}$ have right-inverses with respect to multiplication: for every $r \in \mathbb{F}$ except $r = 0$, there is an element $s \in \mathbb{F}$ such that
$$r \cdot s = 1.$$

7. Multiplication and addition interact via the laws of distributivity:
$$\begin{array}{rcl} (r+s)\cdot t &=& (r\cdot t)+(s\cdot t) \\ t\cdot(r+s) &=& (t\cdot r)+(t\cdot s) \end{array} \qquad \text{for all } r,s,t \in \mathbb{F}.$$

8. Multiplication in $\mathbb{F}$ is commutative:
$$r \cdot s = s \cdot r \qquad \text{for all } r,s \in \mathbb{F}.$$

Remark 2.12. The first three axiom (1-3) just state that $(\mathbb{F}, 0, +)$ is a group. The axioms (4-6) imply that the non-zero elements of $\mathbb{F}$ form a group with $\cdot$ as the law of composition and 1 as identity-element. Hence, what we derived for groups readily applies to fields. In particular we have:

1. The identity-elements 0 and 1 are left-identities, too.

2. There is no other right-identity element with respect to either addition nor multiplication.

3. Every element r has a unique right-inverse with respect to addition, and if $r \neq 0$, it has a unique right-inverse with respect to multiplication.

4. These right-inverses are also left-inverses.

Customarily, the additive inverse of r is denoted by $(-r)$ and the multiplicative inverse of r is denoted by r^{-1}.

Remark 2.13. Some people do not require multiplication to be commutative. They would call our fields "commutative field". I prefer to require commutative fields and call their (defective) fields "skew fields". This disagreement is the reason for listing two forms of the distributive laws. In the presence of commutativity they mutually imply one another.

You may wonder why we include commutativity of multiplication but not commutativity for addition. The reason is that commutativity of addition is actually implied by the other axioms:

Theorem 2.14. *Addition in any field* $\mathbb{F}$ *obeys the law of commutativity:*

$$r + s = s + r \qquad \text{for all } r, s \in \mathbb{F}.$$

Proof. First of all, we define

$$2 := 1 + 1$$

and observe that by axioms 5 and 7:

$$r + r = (r \cdot 1) + (r \cdot 1) = r \cdot (1 + 1) = r \cdot 2 \tag{1}$$

With this lemma, we can carry out the crucial computation:

$$\begin{aligned} (r+r)+(s+s) &= (r \cdot 2) + (s \cdot 2) && \text{by equation 1} \\ &= (r+s) \cdot 2 && \text{by axiom 7} \\ &= (r+s)+(r+s) && \text{by equation 1} \end{aligned}$$

The crucial part is how the order of the two middle terms has changed. We exploit this as follows:

$$\begin{aligned} r+s &= (r+s)+0 && \text{by axiom 2} \\ &= (r+s)+(s+(-s)) && \text{by axiom 3} \\ &= 0+((r+s)+(s+(-s))) && \text{since 0 is also a left-zero} \\ &= ((-r)+r)+((r+s)+(s+(-s))) && \text{since } (-\mathbb{F}) \text{ is also a left-inverse} \\ &= \big((-r)+((r+r)+(s+s))\big)+(-s) && \text{repeated use of axiom 1} \\ &= \big((-r)+((r+s)+(r+s))\big)+(-s) && \text{by our crucial computation} \\ &= s+r && \text{by reversing the steps above} \end{aligned}$$

q.e.d.

Example 2.15. You already know the fields $\mathbb{Q}$ (the set of all rational numbers) and $\mathbb{R}$ (the set of all real numbers). Some of you may also know the set $\mathbb{C}$ of complex numbers. All of these are fields. The integers $\mathbb{Z}$ do not form a field since we cannot divide (most integers lack multiplicative inverses!)

Exercise 2.16. Recall the group $\{\text{even}, \text{odd}\}$ with two elements from Exercise 2.9. Define a multiplication on this set as follows

·	even	odd
even	even	even
odd	even	odd

Show that these conventions turn $\{\text{even}, \text{odd}\}$ into a field with even as its zero and odd as its one.

Remark 2.17. This field with two elements is used in computer science and coding theory. Linear algebra over this peculiar field is actually quite useful.

3 Vector Spaces

3.1 Definition and First Properties

Definition 3.1 (see book 1.1). Let $\mathbb{F}$ be a field. A vector space over $\mathbb{F}$ (or short, an $\mathbb{F}$-vector space) is a set V together with a distinguished element $\mathbf{0}$ [called the zero-vector] and two binary operations

$$\begin{aligned} \oplus : V \times V &\longrightarrow V \\ (\mathbf{v}, \mathbf{w}) &\mapsto \mathbf{v} \oplus \mathbf{w} \end{aligned}$$

[called (vector) addition] and

$$\begin{aligned} \odot : \mathbb{F} \times V &\longrightarrow V \\ (r, \mathbf{v}) &\mapsto r \odot \mathbf{v} \end{aligned}$$

[called (scalar) multiplication] such that the following axioms hold:

1. ~~(commutativity of addition) We have $\mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v}$ for all $\mathbf{v}, \mathbf{w} \in V$.~~
2. (associativity of addition) We have
$$(\mathbf{v} \oplus \mathbf{w}) \oplus \mathbf{x} = \mathbf{v} \oplus (\mathbf{w} \oplus \mathbf{x}) \qquad \text{for all } \mathbf{v}, \mathbf{w}, \mathbf{x} \in V.$$
3. (additive right identity) We have
$$\mathbf{v} \oplus \mathbf{0} = \mathbf{v} \qquad \text{for all } \mathbf{v} \in V.$$
4. (existence of additive right inverses) For each $\mathbf{v} \in V$, there is a vector $\mathbf{x} \in V$ (depending on $\mathbf{v}$) satisfying $\mathbf{v} \oplus \mathbf{x} = \mathbf{0}$.
5. (distributivity I) We have
$$r \odot (\mathbf{v} \oplus \mathbf{w}) = (r \odot \mathbf{v}) \oplus (r \odot \mathbf{w}) \qquad \text{for all } r \in \mathbb{F} \text{ and } \mathbf{v}, \mathbf{w} \in V.$$
6. (distributivity II) We have
$$(r + s) \odot \mathbf{v} = (r \odot \mathbf{v}) \oplus (s \odot \mathbf{v}) \qquad \text{for all } r, s \text{ and } \mathbf{v} \in V.$$
7. (associativity of multiplications) We have
$$(r \cdot s) \odot \mathbf{v} = r \odot (s \odot \mathbf{v}) \qquad \text{for all } r, s \in \mathbb{F} \text{ and } \mathbf{v} \in V.$$
8. (multiplicative identity) We have
$$1 \odot \mathbf{v} = \mathbf{v} \qquad \text{for all } \mathbf{v} \in V.$$

Proposition 3.2 (2–4). *If* $\mathbf{v} \oplus \mathbf{w} = \mathbf{0}$, *then* $\mathbf{w} \oplus \mathbf{v} = \mathbf{0}$.

Proof. First, we chose $\mathbf{x}$ so that $\mathbf{w} \oplus \mathbf{x} = \mathbf{0}$. Such a vector $\mathbf{x}$ exists by Axiom 3.1.4. Now, we compute:

$$\begin{aligned}
\mathbf{w} \oplus \mathbf{v} &= \mathbf{w} \oplus (\mathbf{v} \oplus \mathbf{0}) && \text{by Axiom 2} \\
&= \mathbf{w} \oplus (\mathbf{v} \oplus (\mathbf{w} \oplus \mathbf{x})) && \text{by choice of } \mathbf{x} \\
&= \mathbf{w} \oplus ((\mathbf{v} \oplus \mathbf{w}) \oplus \mathbf{x}) && \text{by Axiom 3.1.2} \\
&= \mathbf{w} \oplus (\mathbf{0} \oplus \mathbf{x}) && \text{since } \mathbf{v} \oplus \mathbf{w} = \mathbf{0} \\
&= (\mathbf{w} \oplus \mathbf{0}) \oplus \mathbf{x} && \text{by Axiom 3.1.2} \\
&= \mathbf{w} \oplus \mathbf{x} && \text{by Axiom 3.1.3} \\
&= \mathbf{0} && \text{by choice of } \mathbf{x}
\end{aligned}$$

q.e.d.

Proposition 3.3 (2–4). *We have*

$$\mathbf{0} \oplus \mathbf{v} = \mathbf{v} \qquad \textit{for all } \mathbf{v} \in V.$$

Proof. Using Axiom 3.1.4, we choose some $\mathbf{x}$ satisfying $\mathbf{v} \oplus \mathbf{x} = \mathbf{0}$. Then:

$$\begin{aligned}
\mathbf{0} \oplus \mathbf{v} &= (\mathbf{v} \oplus \mathbf{x}) \oplus \mathbf{v} && \text{by choice of } \mathbf{x} \\
&= \mathbf{v} \oplus (\mathbf{x} \oplus \mathbf{v}) && \text{by Axiom 3.1.2} \\
&= \mathbf{v} \oplus \mathbf{0} && \mathbf{x} \oplus \mathbf{v} = \mathbf{0} \text{ by Proposition 3.2} \\
&= \mathbf{v} && \text{by Axiom 3.1.3}
\end{aligned}$$

q.e.d.

Proposition 3.4 (2–4). *For all* $\mathbf{v}, \mathbf{w} \in V$*, there is a* <u>*unique*</u> *vector* $\mathbf{x} \in V$ *satisfying*

$$\mathbf{v} \oplus \mathbf{x} = \mathbf{w}.$$

The uniqueness statement means that $\mathbf{v} \oplus \mathbf{x} = \mathbf{w} = \mathbf{v} \oplus \mathbf{y}$ *implies* $\mathbf{x} = \mathbf{y}$*.*

Proof. Using Axiom 3.1.4, we chose $\mathbf{v}'$ so that $\mathbf{v} \oplus \mathbf{v}' = \mathbf{0}$. There are two parts of this statement that need to be argued separately.

existence: Now, we claim that $\mathbf{x} := \mathbf{v}' \oplus \mathbf{w}$ satisfies the equation. To see this, we just plug and chuck:

$$\begin{aligned}
\mathbf{v} \oplus \mathbf{x} &= \mathbf{v} \oplus (\mathbf{v}' \oplus \mathbf{w}) && \text{by choice of } \mathbf{x} \\
&= (\mathbf{v} \oplus \mathbf{v}') \oplus \mathbf{w} && \text{by Axiom 3.1.2} \\
&= \mathbf{0} \oplus \mathbf{w} && \text{by choice of } \mathbf{v}' \\
&= \mathbf{w} && \text{by Proposition 3.3}
\end{aligned}$$

uniqueness: We have to prove that

$$\mathbf{v} \oplus \mathbf{x} = \mathbf{v} \oplus \mathbf{y} \quad \text{implies} \quad \mathbf{x} = \mathbf{y}.$$

We compute

$$\begin{array}{r|l} \mathbf{v} \oplus \mathbf{x} = \mathbf{v} \oplus \mathbf{y} & \text{hypothesis} \\ \mathbf{v}' \oplus (\mathbf{v} \oplus \mathbf{x}) = \mathbf{v}' \oplus (\mathbf{v} \oplus \mathbf{y}) & \text{adding } \mathbf{v}' \\ (\mathbf{v}' \oplus \mathbf{v}) \oplus \mathbf{x} = (\mathbf{v}' \oplus \mathbf{v}) \oplus \mathbf{y} & \text{by Axiom 3.1.2} \\ \mathbf{0} \oplus \mathbf{x} = \mathbf{0} \oplus \mathbf{y} & \mathbf{v}' \oplus \mathbf{v} = \mathbf{0} \text{ by Proposition 3.2} \\ \mathbf{x} = \mathbf{y} & \text{by Proposition 3.3} \end{array}$$

q.e.d.

Proposition 3.5 (2–4, see book 1.8). *For all* $\mathbf{v}, \mathbf{w} \in V$*, there is a* <u>*unique*</u> *vector* $\mathbf{y} \in V$ *satisfying*

$$\mathbf{y} \oplus \mathbf{v} = \mathbf{w}.$$

The uniqueness part means that $\mathbf{x} \oplus \mathbf{v} = \mathbf{w} = \mathbf{y} \oplus \mathbf{v}$ *implies* $\mathbf{x} = \mathbf{y}$*.*

Proof. Proposition 3.2 and Proposition 3.3 are mirror images of Axioms 3 and 4. Thus, Proposition 3.5 holds as it is the mirror image of Proposition 3.4. The real proof is left as an exercise. **q.e.d.**

Corollary 3.6 (2–4, see Theorem 1.3). *For each* $\mathbf{v}$ *there is a unique* $\mathbf{v}$ *with*

$$\mathbf{v} \oplus \mathbf{x} = \mathbf{0}$$

We denote this vector $\mathbf{x}$ *by* $-\mathbf{v}$*. It also satisfies* $-\mathbf{v} \oplus \mathbf{v} = \mathbf{0}$ *and is also uniquely determined by this condition.*

Proof. Use Proposition 3.4 with $\mathbf{w} = \mathbf{0}$. For the symmetric statements, use Proposition 3.5. **q.e.d.**

Proposition 3.7 (see book 1.4). *We have*

$$0 \odot \mathbf{v} = \mathbf{0} \qquad \textit{for all } \mathbf{v} \in V.$$

Proof. First, we compute

$$\begin{array}{rl|l} \mathbf{v} \oplus (0 \odot \mathbf{v}) & = (1 \odot \mathbf{v}) \oplus (0 \odot \mathbf{v}) & \text{by Axiom 3.1.8} \\ & = (1 + 0) \odot \mathbf{v} & \text{Axiom 3.1.6} \\ & = 1 \odot \mathbf{v} & \text{since } 1 + 0 = 1 \\ & = \mathbf{v} & \text{by Axiom 3.1.8} \end{array}$$

Thus, $0 \odot \mathbf{v}$ is a solution to the equation

$$\mathbf{v} \oplus \mathbf{x} = \mathbf{v}.$$

This, equation however has the obvious solution $\mathbf{x} = \mathbf{0}$ by Axiom 3.1.3. Since there is only one solution to this equation (uniqueness part of Proposition 3.4), we find that the two solutions we have must coincide, i.e., $0 \odot \mathbf{v} = \mathbf{0}$. **q.e.d.**

Proposition 3.8 (see book 1.5(d)). *We have*

$$(-1) \odot \mathbf{v} = -\mathbf{v} \qquad \textit{for all } \mathbf{v} \in V.$$

Proof. We have to show that $(-1) \odot \mathbf{v}$ is an additive inverse for $\mathbf{v}$, i.e., we need to show

$$\mathbf{v} \oplus ((-1) \odot \mathbf{v}) = \mathbf{0}.$$

This can be done like so:

$$\begin{aligned} \mathbf{v} \oplus ((-1) \odot \mathbf{v}) &= (1 \odot \mathbf{v}) \oplus ((-1) \odot \mathbf{v}) && \text{by Axiom 3.1.8} \\ &= (1 + (-1)) \odot \mathbf{v} && \text{by Axiom 3.1.6} \\ &= 0 \odot \mathbf{v} && \\ &= \mathbf{0} && \text{by Proposition 3.7} \end{aligned}$$

q.e.d.

Lemma 3.9. *If* $\mathbf{v} \oplus \mathbf{v} = \mathbf{v}$*, then* $\mathbf{v} = \mathbf{0}$.

Proof. Consider the equation

$$\mathbf{v} \oplus \mathbf{x} = \mathbf{v}.$$

By Proposition 3.4, this equation has a unique solution, which by Axiom 3.1.3 must be $\mathbf{x} = \mathbf{0}$ and by hypothesis ($\mathbf{v} \oplus \mathbf{v} = \mathbf{v}$) must be $\mathbf{x} = \mathbf{v}$. **q.e.d.**

Proposition 3.10 (see book 1.5(b)). *We have*

$$r \odot \mathbf{0} = \mathbf{0} \qquad \textit{for all } r \in \mathbb{F}.$$

Proof. Note

$$\begin{aligned} (r \odot \mathbf{0}) \oplus (r \odot \mathbf{0}) &= r \odot (\mathbf{0} \oplus \mathbf{0}) && \text{by Axiom 3.1.5} \\ &= r \odot \mathbf{0} && \text{by Axiom 3.1.3} \end{aligned}$$

Thus, $r \odot \mathbf{0}$ satisfies the hypothesis of Lemma 3.9. Hence $r \odot \mathbf{0} = \mathbf{0}$. **q.e.d.**

Proposition 3.11 (compare book 1.5(c,i)). *If* $r \neq 0$ *and* $r \odot \mathbf{v} = \mathbf{0}$*, then* $\mathbf{v} = \mathbf{0}$.

Proof. Since $r \neq 0$, there is a scalar s such that $r \cdot s = 1$. Since multiplication in $\mathbb{F}$ is commutative, we have $s \cdot r = 1$, as well. (Exercise: find a way to avoid the reference to commutativity of multiplication in $\mathbb{F}$.) Now, we compute:

$$\begin{aligned} \mathbf{v} &= 1 \odot \mathbf{v} && \text{by Axiom 3.1.8} \\ &= (s \cdot r) \odot \mathbf{v} && \text{since } s \cdot r = 1 \\ &= s \odot (r \odot \mathbf{v}) && \text{by Axiom 3.1.7} \\ &= s \odot \mathbf{0} && \text{by hypothesis } r \odot \mathbf{v} = \mathbf{0} \\ &= \mathbf{0} && \text{by Proposition 3.10} \end{aligned}$$

q.e.d.

Theorem 3.12 (commutativity of addition). *We have*

$$\mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v} \qquad \text{for all } \mathbf{v}, \mathbf{w} \in V.$$

Proof. First, we argue

$$(\mathbf{v} \oplus \mathbf{v}) \oplus (\mathbf{w} \oplus \mathbf{w}) = (\mathbf{v} \oplus \mathbf{w}) \oplus (\mathbf{v} \oplus \mathbf{w}) \tag{*}$$

as follows:

$$\begin{aligned}
(\mathbf{v} \oplus \mathbf{v}) \oplus (\mathbf{w} \oplus \mathbf{w}) &= ((1 \odot \mathbf{v}) \oplus (1 \odot \mathbf{v})) \oplus ((1 \odot \mathbf{w}) \oplus (1 \odot \mathbf{w})) && \text{by Axiom 3.1.8} \\
&= ((1+1) \odot \mathbf{v}) \oplus ((1+1) \odot \mathbf{w}) && \text{by Axiom 3.1.6} \\
&= (1+1) \odot (\mathbf{v} \oplus \mathbf{w}) && \text{by Axiom 3.1.5} \\
&= (1 \odot (\mathbf{v} \oplus \mathbf{w})) \oplus (1 \odot (\mathbf{v} \oplus \mathbf{w})) && \text{by Axiom 3.1.6} \\
&= (\mathbf{v} \oplus \mathbf{w}) \oplus (\mathbf{v} \oplus \mathbf{w}) && \text{by Axiom 3.1.8}
\end{aligned}$$

Now that we established (*), we argue:

$$\begin{aligned}
(\mathbf{v} \oplus \mathbf{v}) \oplus (\mathbf{w} \oplus \mathbf{w}) &= (\mathbf{v} \oplus \mathbf{w}) \oplus (\mathbf{v} \oplus \mathbf{w}) && \text{this is (*)} \\
\mathbf{v} \oplus (\mathbf{v} \oplus (\mathbf{w} \oplus \mathbf{w})) &= \mathbf{v} \oplus (\mathbf{w} \oplus (\mathbf{v} \oplus \mathbf{w})) && \text{by Axiom 3.1.2} \\
\mathbf{v} \oplus (\mathbf{w} \oplus \mathbf{w}) &= \mathbf{w} \oplus (\mathbf{v} \oplus \mathbf{w}) && \text{by Proposition 3.4 (uniqueness)} \\
(\mathbf{v} \oplus \mathbf{w}) \oplus \mathbf{w} &= (\mathbf{w} \oplus \mathbf{v}) \oplus \mathbf{w} && \text{by Axiom 3.1.2} \\
\mathbf{v} \oplus \mathbf{w} &= \mathbf{w} \oplus \mathbf{v} && \text{by Proposition 3.5 (uniqueness)}
\end{aligned}$$

q.e.d.

3.2 Examples

Example 3.13 (You can be a vector). The one-element set $V := \{$🯅$\}$ is a vector space when endowed with the following structure:

1. The zero-vector is $\mathbf{0} :=$ 🯅.
2. Addition is defined by 🯅 $\oplus$ 🯅 $:=$ 🯅.
3. Multiplication is defined as $r \odot$ 🯅 $:=$ 🯅 for all $r \in \mathbb{F}$.

Proof. First, note that the only element is its own inverse: 🯅 $\oplus$ 🯅 $=$ 🯅 $= \mathbf{0}$. Since this covers all elements, we have established Axiom 3.1.4. All other axioms just require that some equation LHS = RHS holds in general. However, note that any expression eventually will evaluate to 🯅 as this is the only legitimate value. Thus any equation will hold in this case. In particular all those that are required by the axioms. **q.e.d.**

Example 3.14 (standard vector spaces). We endow the set

$$\mathbb{R}^m := \{(x_1, x_2, \ldots, x_m) \mid x_1, x_2, \ldots, x_m \in \mathbb{R}\}$$

with the following structure

1. The zero-vector is the all-0-tuple: $\mathbf{0} := (0, 0, \ldots, 0)$.
2. Addition is given by $(x_1, \ldots, x_m) \oplus (y_1, \ldots, y_m) := (x_1 + y_1, \ldots, x_m + y_m)$.
3. We define multiplication by $r \odot (x_1, \ldots, x_m) := (rx_1, \ldots, rx_m)$

This way, $\mathbb{R}^m$ is a vector space over $\mathbb{R}$.

Proof. We start by verifying Axiom 3.1.4. Of course, there is but one reasonable candidate for the inverse of $(x_1, \ldots, x_m)$, namely: $(-x_1, \ldots, -x_m)$. Indeed, we check:

$$\begin{aligned}
(x_1, \ldots, x_m) \oplus (-x_1, \ldots, -x_m) &= (x_1 + (-x_1), \ldots, x_m + (-x_m)) && \text{by definition of } \oplus \\
&= (0, \ldots, 0) && \text{by arithmetic in } \mathbb{R} \\
&= \mathbf{0} && \text{by definition of } \mathbf{0} \in \mathbb{R}^m
\end{aligned}$$

The recipe for all other axioms is as follows: To prove an identity in $\mathbb{R}^m$, write out the left hand side eliminating all occurrences of $\oplus$ and $\odot$ using the definitions. Afterwards, use the corresponding rules for real numbers to check that the left hand and the right hand match. Here is a sample demonstration of Axiom 3.1.5: We have to show that

$$r \odot ((x_1, \ldots, x_m) \oplus (y_1, \ldots, y_m)) = (r \odot (x_1, \ldots, x_m)) \oplus (r \odot (y_1, \ldots, y_m))$$

Thus, we evaluate

$$\begin{aligned}
\text{LHS} &:= r \odot ((x_1, \ldots, x_m) \oplus (y_1, \ldots, y_m)) && \\
&= r \odot (x_1 + y_1, \ldots, x_m + y_m) && \text{by definition of } \oplus \\
&= (r(x_1 + y_1), \ldots, r(x_m + y_m)) && \text{by definition of } \odot
\end{aligned}$$

and

$$\begin{aligned}
\text{RHS} &:= (r \odot (x_1, \ldots, x_m)) \oplus (r \odot (y_1, \ldots, y_m)) && \\
&= (rx_1, \ldots, rx_m) \oplus (ry_1, \ldots, ry_m) && \text{by definition of } \odot \\
&= (rx_1 + ry_1, \ldots, rx_m + ry_m) && \text{by definition of } \oplus
\end{aligned}$$

Now, it follows that LHS = RHS since

$$\begin{aligned}
r(x_1 + y_1) &= rx_1 + ry_1 \\
&\vdots \\
r(x_m + y_m) &= rx_m + ry_m
\end{aligned}$$

Note how the law of distributivity for real numbers is used to argue its counter part Axiom 3.1.5. The other axioms are verified in exactly the same fashion. This is left as an exercise. **q.e.d.**

Example 3.15 (perverted operations). We endow the set $V := \mathbb{R}$ with the following (perverted) structure:

1. The zero-vector is $\mathbf{0} := -2$.
2. Addition is given by $\mathbf{v} \oplus \mathbf{w} := \mathbf{v} + \mathbf{w} + 2$.
3. We multiply as follows: $t \odot \mathbf{v} := t\mathbf{v} + 2t - 2$.

These absurd conventions do, in fact, turn V into a vector space over $\mathbb{R}$.

Proof. We shall just demonstrate two axioms: Axiom 3.1.5, and Axiom 3.1.4. For Axiom 3.1.5, we have to show:

$$\text{LHS} := r \odot (\mathbf{v} \oplus \mathbf{w}) = (r \odot \mathbf{v}) \oplus (r \odot \mathbf{w}) =: \text{RHS}.$$

Now, we unravel the definitions (getting rid of circles symbols):

$$\begin{aligned} \text{LHS} &= r \odot (\mathbf{v} \oplus \mathbf{w}) && \text{by defition of LHS} \\ &= r(\mathbf{v} \oplus \mathbf{w}) + 2r - 2 && \text{by defintion of } \odot \\ &= r(\mathbf{v} + \mathbf{w} + 2) + 2r - 2 && \text{by defintion of } \oplus \\ &= r(\mathbf{v} + \mathbf{w} + 4) - 2 && \text{by real arithmetic} \end{aligned}$$

$$\begin{aligned} \text{RHS} &= (r \odot \mathbf{v}) \oplus (r \odot \mathbf{w}) && \text{by definition of RHS} \\ &= (r \odot \mathbf{v}) + (r \odot \mathbf{w}) + 2 && \text{by definition of } \oplus \\ &= (r\mathbf{v} + 2r - 2) + (r\mathbf{w} + 2r - 2) + 2 && \text{by definition of } \odot \\ &= r(\mathbf{v} + \mathbf{w} + 4) - 2 && \text{by real arithmetic} \\ &= \text{LHS} \end{aligned}$$

For Axiom 3.1.4, we have to find additive inverses. Because of Proposition 3.8, there is but one candidate for $-\mathbf{v}$, namely $(-1) \odot \mathbf{v} = -\mathbf{v} - 4$. Now, we actually find:

$$\begin{aligned} \mathbf{v} \oplus (-\mathbf{v} - 4) &= \mathbf{v} + (-\mathbf{v} - 4) + 2 && \text{by definition of } \oplus \\ &= -2 && \text{by real arithmetic} \\ &= \mathbf{0} && \text{by definition of } \mathbf{0} \end{aligned}$$

Thus, each $\mathbf{v}$ has an additive inverse. **q.e.d.**

Example 3.16 (the big function space). Let V be a vector space over the field $\mathbb{F}$, and let M be a set. We endow the set $\text{Maps}(M; V) := \{f : M \to V\}$ of all V-valued functions over the domain M with the following structure:

1. The zero-vector in $\text{Maps}(M; V)$ is the map that is $\mathbf{0}_V$ everywhere:

$$\begin{aligned} \mathbf{0} : M &\longrightarrow V \\ m &\mapsto \mathbf{0}_V \end{aligned}$$

2. Addition of vectors is defined point-wise (the vectors are maps!):

$$f \oplus h := (M \ni m \mapsto f(m) \oplus_V h(m))$$

i.e.,

$$(f \oplus h)(m) := f(m) \oplus_V h(m) \qquad \text{for all } m \in M.$$

3. Multiplication is defined point-wise, too:

$$r \odot f := (M \ni m \mapsto r \odot_V f(m))$$

i.e.,

$$(r \odot f)(m) := r \odot_V (f(m)) \qquad \text{for all } m \in M.$$

(Here a V-subscript indicates that the corresponding symbol belongs to the $\mathbb{F}$ vector space structure of V.) These definitions turn $\mathrm{Maps}(M;V)$ into a vector space over $\mathbb{F}$.

Proof. We argue axiom 3.1.4 first: The inverse for a map f will be the map

$$x \mapsto -f(x)\,.$$

This does the trick, since we have:

$$\begin{aligned} (f \oplus (x \mapsto -f(x)))(m) &= f(m) \oplus_V (-f(m)) \\ &= \mathbf{0}_V \end{aligned} \qquad \Big|$$

Hence

$$(f \oplus (x \mapsto -f(x))) = \mathbf{0}.$$

The other axioms transfer from V in the usual way. We just illustrate Axiom 3.1.6:

$$\mathrm{LHS} := (r+s) \odot f = (r \odot f) \oplus (s \odot f) =: \mathrm{RHS}.$$

We find:

$$\begin{aligned} \mathrm{LHS} &= (r+s) \odot f && \Big| \text{ by definition of LHS} \\ &= (m \mapsto (r+s) \odot_V f(m)) && \Big| \text{ by definition of } \odot \end{aligned}$$

and

$$\begin{aligned} \mathrm{RHS} &= (r \odot f) \oplus (s \odot f) && \Big| \text{ by definition of RHS} \\ &= (m \mapsto (r \odot f)(m) \oplus_V (s \odot f)(m)) && \Big| \text{ by definition of } \oplus \\ &= (m \mapsto (r \odot_V f(m)) \oplus_V (s \odot_V f(m))) && \Big| \text{ by definition of } \odot \\ &= (m \mapsto (r+s) \odot_V f(m)) && \Big| \text{ by Axiom 3.1.6 for } V \\ &= \mathrm{LHS} && \Big| \end{aligned}$$

The other axioms are done exactly the same way. **q.e.d.**

3.3 Subspaces

Definition 3.17. Let $(V, \mathbf{0}, \oplus, \odot)$ be a vector space over the field $\mathbb{F}$. Let S be a subset of V. We say that S is closed w.r.t. addition if

$$\mathbf{v} \oplus \mathbf{w} \in S \text{ for all } \mathbf{v}, \mathbf{w} \in S.$$

We say that S is closed w.r.t. multiplication if

$$r \odot \mathbf{v} \in S \text{ for all } r \in \mathbb{F} \text{ and all } \mathbf{v} \in S.$$

Exercise 3.18. Show that the empty set is closed with respect to addition and multiplication.

Observation 3.19. *If $S \subseteq V$ is closed with respect to addition, then addition in the ambient vector space V (denoted by $\oplus_V$) induces a binary operaton on S (denoted by $\oplus_S$):*

$$\begin{aligned} \oplus_S : S \times S &\longrightarrow S \\ (\mathbf{v}, \mathbf{w}) &\mapsto \mathbf{v} \oplus_V \mathbf{w} \end{aligned}$$

Theorem 3.20 (Subspace Theorem). *Let $(V, \mathbf{0}_V, \oplus_V, \odot_V)$ be a vector space over the field $\mathbb{F}$. Suppose a subset $S \subseteq V$ satisfies the following conditions:*

1. *S is non-empty.*
2. *S is closed with respect to addition.*
3. *S is closed with respect to scalar multiplication.*

Then, the following conventions turn S into a vector space:

$\mathbf{0}_S := \mathbf{0}_V.$

$\mathbf{v} \oplus_S \mathbf{w} := \mathbf{v} \oplus_V \mathbf{w}$ *for all* $\mathbf{v}, \mathbf{w} \in S.$

$r \odot_S \mathbf{v} := r \odot_V \mathbf{v}$ *for all* $r \in \mathbb{F}$ *and* $\mathbf{v} \in S.$

That is, with the above definitions, $(S, \mathbf{0}_S, \oplus_S, \odot_S)$ is an $\mathbb{F}$-vector space.

Proof of the Subspace Theorem 3.20. Since $\mathbf{0}_V = 0 \odot_V \mathbf{v}$, we find that $\mathbf{0}_V \in S$ since S is non-empty and closed with respect to multiplication.

All axioms but axiom 3.1.4 are inherited from their counterparts for $\oplus_V$ and $\odot_V$. Existence of inverses, however, also is easy: By Proposition 3.8 $(-1) \odot_V \mathbf{v}$ is an additive inverse for $\mathbf{v} \in V$. If $\mathbf{v} \in S$, then $(-1) \odot_V \mathbf{v} \in S$ and therefore, we have an inverse in S. **q.e.d.**

Definition 3.21 (Subspace). Let $(V, \mathbf{0}, \oplus, \odot)$ be a vector space over the field $\mathbb{F}$. A subspace of V is a non-empty subset of V that is closed with respect to addition and scalar multiplication. By the previous theorem, any subspace inherits the structure of a vector space from the ambient vector space.

Example 3.22 (Polynomials). For each number $d \geq 0$, the set

$$\mathbb{P}_d := \left\{ p : \mathbb{R} \to \mathbb{R} \;\middle|\; \begin{array}{l} \text{There are } a_0, a_1, \ldots, a_d \in \mathbb{R} \text{ such that} \\ p(x) = a_d x^d + a_{d-1} x^{d-1} + a_1 x^1 + a_0 x^0 \\ \text{for all } x \in \mathbb{R}. \end{array} \right\}$$

is a subspace of $\mathrm{Maps}(\mathbb{R}; \mathbb{R})$.

Proof. By the Subspace Theorem, we have to see that the set of polynomials of degree $\leq d$ is closed with respect to addition and scalar multiplication. That, however, is obvious, e.g.:

$$(x \mapsto 3x + 2) \oplus (x \mapsto 2x + (-1)) = (x \mapsto 5x + 1)$$

and

$$2 \odot (x \mapsto x^2 - x + 2) = (x \mapsto 2x^2 - 2x + 4).$$

q.e.d.

Definition 3.23.

Example and Definition 3.24 (Span). Let $(V, \mathbf{0}, \oplus, \odot)$ be an $\mathbb{F}$-vector space. Given vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$, the span

$$\mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} := \{\mathbf{w} \in V \mid (a_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (a_r \odot \mathbf{v}_r) = \mathbf{w} \text{ for some } a_1, \ldots, a_r \in \mathbb{F}\}$$

is a subspace of V.

An expression of the form $(a_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (a_r \odot \mathbf{v}_r)$ is called a linear combination of the vectors $\mathbf{v}_i$. Thus, the span of the $\mathbf{v}_i$ consists of precisely those vectors that can be realized as a linear combination of the $\mathbf{v}_i$. (Warning: the coefficients used in a linear combination to write $\mathbf{w}$ need not be unique!)

Proof. Note that

$$\mathbf{0} = (0 \odot \mathbf{v}_1) \oplus \cdots \oplus (0 \odot \mathbf{v}_r) \in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}.$$

Also, the span is closed with respect to addition: if

$$\mathbf{w} = (a_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (a_r \odot \mathbf{v}_r) \in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$$

and

$$\mathbf{v} = (b_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (b_r \odot \mathbf{v}_r) \in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\}$$

then

$$\begin{aligned} \mathbf{w} \oplus \mathbf{x} &= ((a_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (a_r \odot \mathbf{v}_r)) \oplus ((b_1 \odot \mathbf{v}_1) \oplus \cdots \oplus (b_r \odot \mathbf{v}_r)) \\ &= ((a_1 \odot \mathbf{v}_1) \oplus (b_1 \odot \mathbf{v}_1)) \oplus \cdots \oplus ((a_r \odot \mathbf{v}_r) \oplus (b_r \odot \mathbf{v}_r)) \\ &= ((a_1 + b_1) \odot \mathbf{v}_1) \oplus \cdots \oplus ((a_r + b_r) \odot \mathbf{v}_r)) \\ &\in \mathrm{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r\} \end{aligned}$$

A similar argument (left as an exercise!) shows that the span is closed with respect to scalar multiplication. **q.e.d.**

Example 3.25 (Solutions to a linear equation). The set

$$S := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \;\middle|\; 5x - 7y + z = 0 \right\}$$

is a subspace of $\mathbb{R}^3$.

We can write S as a span: We have

$$\begin{aligned} & \begin{bmatrix} x \\ y \\ z \end{bmatrix} \in S \\ \iff\quad & 5x - 7y + z = 0 \\ \iff\quad & z = -5x + 7y \\ \iff\quad & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} x \\ y \\ -5x + 7y \end{bmatrix} \\ \iff\quad & \begin{bmatrix} x \\ y \\ z \end{bmatrix} = x \begin{bmatrix} 1 \\ 0 \\ -5 \end{bmatrix} + y \begin{bmatrix} 0 \\ 1 \\ 7 \end{bmatrix} \end{aligned}$$

Thus:

$$S = \operatorname{span}\{ \begin{bmatrix} 1 \\ 0 \\ -5 \end{bmatrix}, \begin{bmatrix} 0 \\ 1 \\ 7 \end{bmatrix} \}$$

Morale: solving a linear equation is to write its space of solutions as a span!

4 Linear Maps

4.1 Definition and Examples

Definition 4.1 (Linear Maps). Let V and W be vector spaces. A map $\varphi : V \to W$ is called linear if it satisfies the following two conditions:

1. $\varphi(\mathbf{v} \oplus_V \mathbf{w}) = \varphi(\mathbf{v}) \oplus_W \varphi(\mathbf{w})$ for all $\mathbf{v}, \mathbf{w} \in V$.
2. $\varphi(r \odot_V \mathbf{v}) = r \odot_W \varphi(\mathbf{v})$ for all $r \in \mathbb{F}$ and $\mathbf{v} \in V$.

Remark 4.2. If $\varphi : V \to W$ is linear, then

$$\begin{aligned} \varphi(\mathbf{0}_V) &= \varphi(0 \odot_V \mathbf{0}_V) && \text{by Proposition 3.7} \\ &= 0 \odot_W \varphi(\mathbf{0}_V) && \text{by linearity} \\ &= \mathbf{0}_W && \text{by Proposition 3.7} \end{aligned}$$

Example 4.3 (The derivative). The derivative

$$\begin{aligned} \mathrm{D} : \mathbb{P}_3 &\longrightarrow \mathbb{P}_2 \\ (x \mapsto p(x)) &\mapsto \frac{\mathrm{d}\,p}{\mathrm{d}\,x} \end{aligned}$$

is a linear map.

Proof. By Calculus,

$$\frac{\mathrm{d}(p+q)}{\mathrm{d}\,x} = \frac{\mathrm{d}\,p}{\mathrm{d}\,x} + \frac{\mathrm{d}\,q}{\mathrm{d}\,x}$$

and

$$\frac{\mathrm{d}(rp)}{\mathrm{d}\,x} = r\frac{\mathrm{d}\,p}{\mathrm{d}\,x}.$$

Now, we also observe that the derivative of a polynomial is a polynomial. The dergree drops by 1. **q.e.d.**

Example 4.4 (The evaluation map). Let M be a set and V be a vector space. Fix an elemen $m \in M$. The evaluation at m

$$\begin{aligned} \mathrm{ev}_m : \mathrm{Maps}(M; V) &\longrightarrow V \\ f &\mapsto f(m) \end{aligned}$$

is linear.

Proof. We have:

$$\begin{aligned} \mathrm{ev}_m(f \oplus h) &= (f \oplus h)(m) && \text{definition of } \mathrm{ev}_m \\ &= f(m) \oplus_V h(m) && \text{definition of } \oplus_V \\ &= \mathrm{ev}_m(f) \oplus_V \mathrm{ev}_m(h) && \text{definition of } \mathrm{ev}_m \end{aligned}$$

and:

$$\begin{aligned} \mathrm{ev}_m(r \odot f) &= (r \odot f)(m) && \text{definition of } \mathrm{ev}_m \\ &= r \odot_V f(m) && \text{definition of } \odot_V \\ &= r \odot_V \mathrm{ev}_m(f) && \text{definition of } \mathrm{ev}_m \end{aligned}$$

q.e.d.

Example 4.5 (Spaces of Linear Maps). Let V and W be vector spaces. Then

$$\mathrm{Lin}\,(V;W) := \{\varphi : V \to W \mid \varphi \text{ is linear}\}$$

is a subspace of $\mathrm{Maps}(V;W)$.

Proof. Observe that the zero-map

$$\mathbf{v} \mapsto \mathbf{0}_W \qquad \text{for all } \mathbf{v} \in V$$

is linear. Thus, by the Subspace Theorem, all we have to show is that the sum of two linear maps is linear ($\mathrm{Lin}\,(V;W)$ is closed with respect to addition) and that scalar multiples of linear maps are linear ($\mathrm{Lin}\,(V;W)$ is closed with respect to scalar multiplication).

Now, assume that $\varphi : V \to W$ and $\psi : V \to W$ are linear. Then for any two vector $\mathbf{v}, \mathbf{w} \in V$,

$$\begin{aligned}
(\varphi \oplus \psi)(\mathbf{v} \oplus_V \mathbf{w}) &= \varphi(\mathbf{v} \oplus_V \mathbf{w}) \oplus_W \psi(\mathbf{v} \oplus_V \mathbf{w}) && \text{definition of } \oplus \\
&= (\varphi(\mathbf{v}) \oplus_W \varphi(\mathbf{w})) \oplus_W (\psi(\mathbf{v}) \oplus_W \psi(\mathbf{w})) && \varphi \text{ and } \psi \text{ are linear} \\
&= (\varphi(\mathbf{v}) \oplus_W \psi(\mathbf{v})) \oplus_W (\varphi(\mathbf{w}) \oplus_W \psi(\mathbf{w})) && \text{vector arithmetic in } W \\
&= (\varphi \oplus \psi)(\mathbf{v}) \oplus_W (\varphi \oplus \psi)(\mathbf{w}) && \text{definition of } \oplus
\end{aligned}$$

Hence $\varphi \oplus \psi$ is additive. We also have

$$\begin{aligned}
(\varphi \oplus \psi)(r \odot_V \mathbf{v}) &= \varphi(r \odot_V \mathbf{v}) \oplus_W \psi(r \odot_V \mathbf{v}) && \text{definition of } \oplus \\
&= (r \odot_W \varphi(\mathbf{v})) \oplus_W (r \odot_W \psi(\mathbf{v})) && \varphi \text{ and } \psi \text{ are linear} \\
&= r \odot_W (\varphi(\mathbf{v}) \oplus_W \psi(\mathbf{v})) && \text{vector arithmetic in } W \\
&= r \odot_W ((\varphi \oplus \psi)(\mathbf{v})) && \text{definition of } \oplus
\end{aligned}$$

Thus, $\varphi \oplus \psi$ is linear.

Similar computations show that $s \odot \varphi$ is linear:

$$\begin{aligned}
(s \odot \varphi)(\mathbf{v} \oplus_V \mathbf{w}) &= s \odot_W \varphi(\mathbf{v} \oplus_V \mathbf{w}) && \text{definition of } \odot \\
&= s \odot_W (\varphi(\mathbf{v}) \oplus_W \varphi(\mathbf{w})) && \varphi \text{ is linear} \\
&= (s \odot_W \varphi(\mathbf{v})) \oplus_W (s \odot_W \varphi(\mathbf{w})) && \text{vector arithmetic in } W \\
&= (s \odot \varphi)(\mathbf{v}) \oplus_W (s \odot \varphi)(\mathbf{w}) && \text{definition of } \odot
\end{aligned}$$

and:

$$\begin{aligned}
(s \odot \varphi)(r \odot_V \mathbf{v}) &= s \odot_W \varphi(r \odot_V \mathbf{v}) && \text{definition of } \odot \\
&= s \odot_W (r \odot_W \varphi(\mathbf{v})) && \varphi \text{ is linear} \\
&= r \odot_W (s \odot_W \varphi(\mathbf{v})) && \text{vector arithmetic in } W \\
&= r \odot_W (s \odot \varphi)(\mathbf{v}) && \text{definition of } \odot
\end{aligned}$$

Note how these proofs differ only in the third step: the required "vector arithmetic" varies. You are strongly encouraged to expand this step into an explicit sequence elementary operations. **q.e.d.**

4.2 Linear Maps between Standard Vector Spaces

In this section, we fix a field $\mathbb{F}$ (think $\mathbb{R}$ if you want); and we shall study linear maps from $\mathbb{F}^n$ to $\mathbb{F}^m$ (see example 3.14).

Example 4.6. The map

$$\begin{aligned} \varphi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ \begin{bmatrix} x \\ y \end{bmatrix} &\mapsto -5x + 3y \end{aligned}$$

is linear. In fact, any two scalars $a_1, a_2 \in \mathbb{R}$ define a linear map

$$\begin{aligned} \psi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\ \begin{bmatrix} x \\ y \end{bmatrix} &\mapsto a_1 x + a_2 y. \end{aligned}$$

Moreover, any linear map

$$\psi : \mathbb{R}^2 \longrightarrow \mathbb{R}$$

is of this form. In fact, you can recover the two scalars as

$$\begin{aligned} a_1 &= \psi\left(\begin{bmatrix} 1 \\ 0 \end{bmatrix}\right) \\ a_2 &= \psi\left(\begin{bmatrix} 0 \\ 1 \end{bmatrix}\right) \end{aligned}$$

Proposition 4.7. *Given scalars $a_{i,j} \in \mathbb{F}$, for $i \in \{1, \dots, m\}$ and $j \in \{1, \dots, n\}$, the map*

$$\begin{aligned} \varphi : \mathbb{F}^n &\longrightarrow \mathbb{F}^m \\ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} &\mapsto \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n \end{bmatrix} \end{aligned}$$

is linear.

Proof. This is a straight forward computation. The only difficulty is to not get lost in all those subscripts. The following computation shows that φ is compatible with addition:

$$\begin{aligned} \varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) &= \varphi\left(\begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_n + y_n \end{bmatrix}\right) && \text{addition in } \mathbb{F}^n \\ &= \begin{bmatrix} \sum_{j=1}^n a_{1,j}(x_j + y_j) \\ \vdots \\ \sum_{j=1}^n a_{m,j}(x_j + y_j) \end{bmatrix} && \text{definiton of } \varphi \\ &= \begin{bmatrix} \left(\sum_{j=1}^n a_{1,j}x_j\right) + \left(\sum_{j=1}^n a_{1,j}y_j\right) \\ \vdots \\ \left(\sum_{j=1}^n a_{m,j}x_j\right) + \left(\sum_{j=1}^n a_{m,j}y_j\right) \end{bmatrix} && \text{arithmetic in } \mathbb{F} \\ &= \begin{bmatrix} \sum_{j=1}^n a_{1,j}x_j \\ \vdots \\ \sum_{j=1}^n a_{m,j}x_j \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^n a_{1,j}y_j \\ \vdots \\ \sum_{j=1}^n a_{m,j}y_j \end{bmatrix} && \text{addition in } \mathbb{F}^m \\ &= \varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) + \varphi\left(\begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix}\right) && \text{definition of } \varphi \end{aligned}$$

The computation that shows φ to be compatible with scalar multiplication is similar and left as an exercise. **q.e.d.**

Exercise 4.8. Complete the proof of Proposition 4.7 by showing that φ is compatible with scalar multiplication.

Lemma 4.9. *Let $\varphi : \mathbb{F}^n \to \mathbb{F}^m$ and $\psi : \mathbb{F}^n \to \mathbb{F}^m$ be two linear maps. Suppose that φ and ψ coincide on the "standard basis vectors"*

$$e_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad e_2 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad \cdots \quad e_{n-1} := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad e_n := \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ 1 \end{bmatrix}$$

i.e., assume that $\varphi(e_j) = \psi(e_j)$ *for each* $j \in \{1, \ldots, n\}$.

Then, $\varphi = \psi$*, i.e., for any* $\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \in \mathbb{F}^n$ *we have* $\varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) = \psi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right)$.

Proof. The main observation is

$$\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} = x_1 e_1 + \cdots + x_n e_n$$

Hence, using linearity, we compute:

$$\begin{aligned}
\varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) &= \varphi(x_1 e_1 + \cdots + x_n e_n) && \text{main observation} \\
&= x_1 \varphi(e_1) + \cdots + x_n \varphi(e_n) && \varphi \text{ is linear} \\
&= x_1 \psi(e_1) + \cdots + x_n \psi(e_n) && \varphi \text{ and } \psi \text{ coincide on } e_j \\
&= \psi(x_1 e_1 + \cdots + x_n e_n) && \psi \text{ is linear} \\
&= \psi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) && \text{main observation}
\end{aligned}$$

q.e.d.

Proposition 4.10. *Let*

$$\varphi : \mathbb{F}^n \longrightarrow \mathbb{F}^m$$

be linear. Then there exist uniquely determined scalars $a_{1,1}, \ldots, a_{1,n}, a_{2,1}, \ldots, a_{m,n} \in \mathbb{F}$ *such that*

$$\varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}\right) = \begin{bmatrix} a_{1,1}x_1 + a_{1,2}x_2 + \cdots + a_{1,n}x_n \\ \vdots \\ a_{m,1}x_1 + a_{m,2}x_2 + \cdots + a_{m,n}x_n \end{bmatrix} \text{ for all } \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix} \in \mathbb{F}^m.$$

Proof. We first observe uniqueness: the scalar can be recovered by evaluating φ on the standard basis vectors:

$$\begin{bmatrix} a_{1,1} \\ a_{2,1} \\ \vdots \\ a_{m,1} \end{bmatrix} = \varphi\left(\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right)$$

$$\begin{bmatrix} a_{1,2} \\ a_{2,2} \\ \vdots \\ a_{m,2} \end{bmatrix} = \varphi\left(\begin{bmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right)$$

and so on.

Now let ψ be the linear map defined by the scalars $a_{i,j}$ according to Proposition 4.7. It follows immediately from Lemma 4.9 that $\varphi = \psi$. Thus, φ can be described in the promised way. **q.e.d.**

Customarily, the scalars $a_{i,j}$ are arranged as an $m \times n$-matrix

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \in \mathbb{M}_{m\times n}(\mathbb{F})$$

This allows us to interpret Propositions 4.7 and 4.10 as follows: Proposition 4.7 defines a map

$$\Phi : \mathbb{M}_{m\times n}(\mathbb{F}) \longrightarrow \mathrm{Lin}\left(\mathbb{F}^n; \mathbb{F}^m\right)$$

and Proposition 4.10 defines the inverse

$$\Phi : \mathrm{Lin}\left(\mathbb{F}^n; \mathbb{F}^m\right) \longrightarrow \mathbb{M}_{m\times n}(\mathbb{F}) \,.$$

Recall that $\mathbb{M}_{m\times n}(\mathbb{F})$ and $\mathrm{Lin}\left(\mathbb{F}^n; \mathbb{F}^m\right)$ are vector spaces.

Proposition 4.11. Φ *is linear.*

Proof. We argue compatibility with addition. Let us fix to matrices

$$\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix}, \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ & \vdots & \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix} \in \mathbb{M}_{m\times n}(\mathbb{F}) \,.$$

Put

$$\begin{aligned} \varphi &:= \Phi\left(\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} + \begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ & \vdots & \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix} \right) \\ \psi_1 &:= \Phi\left(\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ & \vdots & \\ a_{m,1} & \cdots & a_{m,n} \end{pmatrix} \right) \\ \psi_2 &:= \Phi\left(\begin{pmatrix} b_{1,1} & \cdots & b_{1,n} \\ & \vdots & \\ b_{m,1} & \cdots & b_{m,n} \end{pmatrix} \right) \end{aligned}$$

We need to verify that $\varphi = \psi_1 + \psi_2$ which is an identity of functions saying

$$\varphi(\mathbf{v}) = (\psi_1 + \psi_2)(\mathbf{v}) = \psi_1(\mathbf{v}) + \psi_2(\mathbf{v}) \quad \text{for each } \mathbf{v} \in \mathbb{F}^n.$$

We compute

$$
\begin{aligned}
\varphi\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) &= \begin{bmatrix} \sum_{j=1}^{n} (a_{1,j} + b_{1,j}) x_j \\ \vdots \\ \sum_{j=1}^{n} (a_{m,j} + b_{m,j}) x_j \end{bmatrix} && \text{by definition of } \varphi \\
&= \begin{bmatrix} \sum_{j=1}^{n} a_{1,j} x_j \\ \vdots \\ \sum_{j=1}^{n} a_{m,j} x_j \end{bmatrix} + \begin{bmatrix} \sum_{j=1}^{n} b_{1,j} x_j \\ \vdots \\ \sum_{j=1}^{n} b_{m,j} x_j \end{bmatrix} && \text{vector arithmetic in } \mathbb{F}^m \\
&= \psi_1\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) + \psi_2\left(\begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}\right) && \text{by definition of } \psi_i
\end{aligned}
$$

The argument that Φ is compatible with scalard multiplication is similar and left as an exercise. **q.e.d.**

Exercise 4.12. Show that Φ is compatible with scalar multiplication.

Exercise 4.13. Show that Ψ is linear.

Example 4.14. The following map is not linear:

$$
\begin{aligned}
\varphi : \mathbb{R}^2 &\longrightarrow \mathbb{R} \\
\begin{bmatrix} x \\ y \end{bmatrix} &\mapsto x^2 - y
\end{aligned}
$$

Proof. This map cannot be described by a matrix in the way of Proposition 4.7. Thus, it cannot be linear by Proposition 4.10.

(Alternatively, one can just give counterexamples to additivity and multiplicativity. That isn't hard.) **q.e.d.**

4.3 Kernel and Image of a Linear Map

Proposition and Definition 4.15. *Let $\mathbb{F}$ be a field, let $(V, \mathbf{0}_V, \oplus_V, \odot_V)$ and $(W, \mathbf{0}_W, \oplus_W, \odot_W)$ be $\mathbb{F}$-vector spaces. Let $\varphi : V \to W$ be a linear map. Then the* <u>*kernel*</u>

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

is a subspace of V.

Proof. By Remark 4.2, we have $\varphi(\mathbf{0}_V) = \mathbf{0}_W$. Thus, $\mathbf{0}_V \in \ker(\varphi)$.

Let $\mathbf{v}_1, \mathbf{v}_2 \in \ker(\varphi)$. Then

$$\begin{aligned} \varphi(\mathbf{v}_1 \oplus_V \mathbf{v}_2) &= \varphi(\mathbf{v}_1) \oplus_W \varphi(\mathbf{v}_2) && \varphi \text{ is linear} \\ &= \mathbf{0}_W \oplus_W \mathbf{0}_W && \mathbf{v}_i \in \ker(\varphi) \\ &= \mathbf{0}_W \end{aligned}$$

Hence, $\ker(\varphi)$ is closed with respect to addition.

Similarly, for $\mathbf{v} \in \ker(\varphi)$ and $r \in \mathbb{F}$, we have

$$\begin{aligned} \varphi(r \odot_V \mathbf{v}) &= r \odot_W \varphi(\mathbf{v}) && \varphi \text{ is linear} \\ &= r \odot_W \mathbf{0}_W && \mathbf{v} \in \ker(\varphi) \\ &= \mathbf{0}_W && \text{by Proposition 3.10} \end{aligned}$$

Hence, $\ker(\varphi)$ is closed with respect to scalar multiplication. **q.e.d.**

Proposition and Definition 4.16. *Let $\mathbb{F}$ be a field, let $(V, \mathbf{0}_V, \oplus_V, \odot_V)$ and $(W, \mathbf{0}_W, \oplus_W, \odot_W)$ be $\mathbb{F}$-vector spaces. Let $\varphi : V \to W$ be a linear map. Then the* <u>*image*</u>

$$\operatorname{im}(\varphi) := \{\mathbf{w} \in W \mid \varphi(\mathbf{v}) = \mathbf{w} \text{ for some } \mathbf{v} \in V\}$$

is a subspace of V.

Proof. Remark 4.2, we have $\varphi(\mathbf{0}_V) = \mathbf{0}_W$. Thus, $\mathbf{0}_W \in \operatorname{im}(\varphi)$.

Let $\mathbf{w}_1, \mathbf{w}_2 \in \operatorname{im}(\varphi)$, i.e., assume that there are vectors $\mathbf{v}_1, \mathbf{v}_2 \in V$ with $\varphi(\mathbf{v}_1) = \mathbf{w}_1$ and $\varphi(\mathbf{v}_2) = \mathbf{w}_2$. Then

$$\begin{aligned} \mathbf{w}_1 \oplus_W \mathbf{w}_2 &= \varphi(\mathbf{v}_1) \oplus_W \varphi(\mathbf{v}_2) && \text{by choice of } \mathbf{v}_1 \text{ and } \mathbf{v}_2 \\ &= \varphi(\mathbf{v}_1 \oplus_V \mathbf{v}_2) && \varphi \text{ is linear} \\ &\in \operatorname{im}(\varphi) && \text{definition of the image} \end{aligned}$$

Hence, $\operatorname{im}(\varphi)$ is closed with respect to addition.

Similarly, let $\mathbf{w} \in \operatorname{im}(\varphi)$ and let $r \in \mathbb{F}$. Thus there is $\mathbf{v} \in V$ with $\varphi(\mathbf{v}) = \mathbf{w}$. Then

$$\begin{aligned} r \odot_W \mathbf{w} &= r \odot_W \varphi(\mathbf{v}) && \text{by choice of } \mathbf{v} \\ &= \varphi(r \odot_V \mathbf{v}) && \varphi \text{ is linear} \\ &\in \operatorname{im}(\varphi) && \text{definition of the image} \end{aligned}$$

Hence, $\operatorname{im}(\varphi)$ is closed with respect to scalar multiplication. **q.e.d.**

5 Linear Systems

5.1 Solving Linear Systems: Preliminary Observations

Example 5.1. Let us consider the linear system

$$\mathcal{L}_1 := \left\| \begin{array}{rrrl} x & +3y & -z & = 9 \\ 2x & +y & +3z & = 8 \end{array} \right\|$$

The vertical bars indicate that both equations are supposed to hold simultaneously. The set of solutions for $\mathcal{L}_1$, therefore, is:

$$S_1 := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \;\middle|\; x + 3y - z = 9 \text{ and } 2x + y + 3z = 8 \right\} \subseteq \mathbb{R}^3.$$

Note that a solution is a triple of numbers and the set of all solutions is a subset of $\mathbb{R}^3$. That set maybe empty (no solutions), a point (a single solution), or infinite. It is this last possibility that requires a little care.

To solve $\mathcal{L}_1$, we first observe that subtracting twice the first row from the second row, yields:

$$\mathcal{L}_2 := \left\| \begin{array}{rrl} -5y & +5z & = -10 \end{array} \right\|$$

This system has the solution set

$$S_2 := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \;\middle|\; -5y + 5z = -10 \right\}.$$

Since the equations of the first system imply the equation in system $\mathcal{L}_2$, we conclude that any solution to $\mathcal{L}_1$ is also a solution to $\mathcal{L}_2$, formally:

$$S_1 \subseteq S_2.$$

The problem is that we might have (and in this case, indeed, have) increased the set of solution: For instance $x = 100, y = 2, z = 0$ solves the second system, but it does not solve the first. Thus: just drawing inferences, we run into the risk of picking up bogus solutions or shadow solutions.

Example 5.2. In high school algebra, the problem of bogus solutions also arises. Consider:

$$\begin{array}{rl|l} \sqrt{x} & = x - 2 & \text{square both sides} \\ x & = x^2 - 4x + 4 & -x \\ 0 & = x^2 - 5x + 4 & \text{factor} \\ 0 & = (x-4)(x-1) & \\ x & \in \{1, 4\} & \end{array}$$

Now, the candidate $x = 1$ does not solve the original equation. Thus, we have to check all candidates and throw those out that to not work. This elimination process is not feasible if we might pick up infinitely many bogus solutions.

Example 5.3. Let us be more careful about drawing our conclusions. We want to replace system $\mathcal{L}_1$ by a somewhat simpler system that has the same set of solutions. Thus, we may not throw away information. Thus, when drawing a conclusion, we also copy some of the original equations. In our example, after subtracting twice the first row from the second, we also copy the first row. Now, we get:

$$\mathcal{L}_3 := \left\| \begin{array}{rrrl} x & +3y & -z & = 9 \\ & -5 & +5z & = -10 \end{array} \right\|$$

The corresponding set of solutions is

$$S_1 := \left\{ \begin{bmatrix} x \\ y \\ z \end{bmatrix} \;\middle|\; x + 3y - z = 9 \text{ and } -5y + 5z = -10 \right\}.$$

I claim:

$$S_1 = S_3.$$

Note that the inclusion $S_1 \subseteq S_3$ follows by the same argument as above. To see that $S_3 \subseteq S_1$ we have to see that the system $\mathcal{L}_3$ implies system $\mathcal{L}_1$. In this case, the two systems mutually imply one another, they are equivalent.

To that $\mathcal{L}_3$ implies $\mathcal{L}_1$, note that adding twice the first row in $\mathcal{L}_3$ to its second row yields the second equation of the first system.

From these considerations, we arrive at the following first crude method:

Rule 5.4. *To solve a linear system, replace the system by a simpler equivalent system. Continue until you cannot simplify any further. Be sure to check at each step that the simplified system and the previous system imply mutually one another, i.e., you have to find an algebraic passage both ways.*

5.2 The Method of Row-Reduction

5.2.1 The Matrix Associated with a Linear System

Given a linear system with r equations in s variables, we can use an $r \times (s+1)$-matrix to encode it. The matrix just encodes the coefficients, and its right column records the right hand sides of the equations.

Example 5.5. Recall our linear system from above:

$$\mathcal{L}_1 := \left\| \begin{array}{rrrl} x & +3y & -z & = 9 \\ 2x & +y & +3z & = 8 \end{array} \right\|$$

The corresponding matrix is

$$A_1 := \begin{bmatrix} 1 & 3 & -1 & 9 \\ 2 & 1 & 3 & 8 \end{bmatrix}$$

5.2.2 The (Reduced) Row-Echelon Form

In this section, we shall make the notion of "solving a linear system" a bit more precise. In our crude wording, we wanted an equivalent system that cannot be further simplified. What that means, however, is not very clear. Passing to matrices allows us to make a more precise definition.

Definition 5.6. Let A be an $m \times n$-matrix.

A leading entry of A is non-zero entry that is the first, from the left, non-zero entry in its row. A leading 1 is a leading entry that happens to be 1.

We say that A is in row-echelon form if it satisfies all of the following three conditions:

(a) Each leading entry is a leading 1.

(b) These leading 1s form a staircase pattern, i.e., given two rows with leading 1s, then the leading 1 in the higher row is strictly to the left of the leading entry in the lower row. (In particular, no column contains two leading 1s.)

(c) All the rows without leading 1s are together at the bottom of A. Note that these rows consist of 0-entries only.

We say that A is in reduced row-echelon form if in addition the following holds:

(d) Above any leading 1 there are only 0-entries in its column.

We consider a linear system solved when its matrix is in reduced row-echelon form.

Example 5.7. The system

$$\left\| \begin{array}{rrl} x & -2z & = 1 \\ y & +3z & = 4 \end{array} \right\|$$

corresponds to

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

which is in reduced row-echelon form. Note how that enables us to solve for x and y in terms of z. This is the key point of the row echelon form: you always can solve for the more left variables in terms of those more to the right.

Example 5.8. The matrix

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

is in reduced row-echelon form. The corresponding system is

$$\left\| \begin{array}{rrrl} 0x & +1y & +2z & = 0 \\ 0x & +0y & -0z & = 1 \end{array} \right\|$$

You can see right away that there are no solutions because of the second equation.

5.2.3 Elementary Row Operations

Definition 5.9. An elementary row operation on a matrix is one of the following three types:

1. The swap interchanges two rows R_i and R_j. We denote the swap by: $R_i \leftrightarrow R_j$.
2. Adding a multiple of some row R_i to some other row R_j. Here we require $i \neq j$. We denote this type by: $CR_i + R_j \to R_j$.
3. Rescaling a row R_j by a non-zero factor $C \neq 0$. This is denoted by: $CR_j \to R_j$.

Observation 5.10. *Let A be a matrix encoding a linear system $\mathcal{L}$. Suppose you obtain a matrix A' from A by an elementary row-operation. Then the system $\mathcal{L}'$ encoded by A' follows from the system $\mathcal{L}$. In particular, any solution to $\mathcal{L}$ is a solution to $\mathcal{L}'$.*

Proposition 5.11. *Every elementary row-operation can be un-done by another elementary row-operation.*

Proof. We have to see, how the three different types of elementary row-operations can be un-done.

As for swaps, this is easy as swaps are self-reversing: swapping twice amounts to doing nothing.

Rescaling by a non-zero scalar (i.e., an operation like $CR_j \to R_j$ with $C \neq 0$) is also easy to un-do: just rescale by $\frac{1}{C}$. (Here, we use the fact that fields have multiplicative inverses for non-zero elements!)

Finally, a modification $CR_i + R_j \to R_j$ is un-done by $-CR_i + R_j \to R_j$. Note: this argument uses $i \neq j$. (Can you see why?) **q.e.d.**

Corollary 5.12. *Let A be a matrix encoding a linear system $\mathcal{L}$. Suppose you obtain a matrix A' from A by an elementary row-operation. Then the system $\mathcal{L}'$ encoded by A' is equivalent to the system $\mathcal{L}$. In particular, both systems have exactly the same set of solutions.*

Remark 5.13. It is important to note that

$$CR_i + R_j \rightarrow R_i$$

is not an elementary row-operation. (Why is this not an elementary row operation? and why is this important? Hint: consider $C = 0$.)

Definition 5.14. A matrices A_1 is equivalent to the matrix A_2 if there is a sequence of elementary row operations fransforming A_1 into matrix A_2.

Remark 5.15. If a matrix describes a linear system, then any equivalent matrix describes a system with the same solution set.

Exercise 5.16. Prove the following:

1. Every matrix is equivalent to itself. (Hint: You have to show that doing nothing can be realised as a sequence of elementary row-operations.)

2. If A_1 is equivalent to A_2, then A_2 is equivalent to A_1. (Hint: You need to show that very sequence of elementary row operations can be un-done by another sequence of elementary row-operations. For a single row-operation, this un-doability is the contents of Proposition 5.11

3. If A_1 is equivalent to A_2 and A_2 is equivalent to A_3, then A_1 is equivalent to A_3. (Hint: Consider concatenation of sequences of row-operations.)

5.2.4 The Gauss-Jordan Algorithm of Row-Reduction

Theorem 5.17 (existence of the reduced row-echelon form). *Every matrix A is equivalent to a matrix in reduced row-echelon form.*

Proof by giving an algorithm. We have to find a sequence of elementary row-operations that puts A into reduced row-echelon form. The following method does not try to be particularly smart, but it will always succeed:

1. If needed, swap two rows so that the first row has a left-most leading entry, i.e., no other row shall have a leading entry farther to the left.

2. If needed, rescale the first row so that its leading entry becomes 1.

3. Subtract multiples of the first row from all other rows to make the leading 1 in the first row the only non-zero entry in its column.

4. Restrict your attention to the submatrix of all rows below the first row, and restart.

5. Continue the loop of the first four steps until the submatrix under consideration has no non-zero entries. Note: the empty matrix has no entries, hence it has no non-zero entries.

6. Now, the matrix is in row-echelon form. To promote it to reduced row-echelon form, for each row with a leading 1 add multiples of that row to all higher rows as to kill the entries above the leading 1.

The inner workings of this method are best understood by working through a (not so random) example: Here is the matrix:

$$\begin{bmatrix} 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 3 & 4 & 5 & 0 & 2 & -1 \\ 6 & 7 & 8 & 1 & 1 & -3 \\ 6 & 7 & 8 & 3 & 2 & 0 \\ 3 & 4 & 5 & 2 & 3 & 2 \end{bmatrix}$$

Step 1 says, we have to swap $R_1 \leftrightarrow R_3$ which yields:

$$\begin{bmatrix} 3 & 4 & 5 & 0 & 2 & -1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 6 & 7 & 8 & 1 & 1 & -3 \\ 6 & 7 & 8 & 3 & 2 & 0 \\ 3 & 4 & 5 & 2 & 3 & 2 \end{bmatrix}$$

Step 2 changes the leading entry in the top row to 1 by $1/3R_1 \to R_1$. We obtain:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 6 & 7 & 8 & 1 & 1 & -3 \\ 6 & 7 & 8 & 3 & 2 & 0 \\ 3 & 4 & 5 & 2 & 3 & 2 \end{bmatrix}$$

Step 3 is about killing the entries below the first leading 1.
$-6R_1 + R_4 \to R_4$ yields

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & -1 & -2 & 1 & -3 & -1 \\ 6 & 7 & 8 & 3 & 2 & 0 \\ 3 & 4 & 5 & 2 & 3 & 2 \end{bmatrix}$$

$-6R_1 + R_5 \to R_5$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & -1 & -2 & 1 & -3 & -1 \\ 0 & -1 & -2 & 3 & -2 & 2 \\ 3 & 4 & 5 & 2 & 3 & 2 \end{bmatrix}$$

$-3R_1 + R_6 \to R_6$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & -1 & -2 & 1 & -3 & -1 \\ 0 & -1 & -2 & 3 & -2 & 2 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

Now, the first column is fixed. We discard the first row and restart the loop.
$R_2 \leftrightarrow R_3$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & -1 & -2 & 1 & -3 & -1 \\ 0 & -1 & -2 & 3 & -2 & 2 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

$1R_2 + R_4 \to R_4$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -1 & -2 & 3 & -2 & 2 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

$1R_2 + R_5 \to R_5$ yields

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

Now, the second row is taken care of. In the next loop, we are lucky and can skip step 1.

$1/2R_3 \to R_3$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

$-2R_3 + R_5 \to R_5$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 3 \end{bmatrix}$$

$-2R_3 + R_6 \to R_6$ yields:

$$\begin{bmatrix} 1 & 4/3 & 5/3 & 0 & 2/3 & -1/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

We finished the third row, and oops: all bottom rows have died. Thus, we stop. Note that we achieved row echelon form.

Finally, we promote this to reduced row-echelon form.

$-4/3R_2 + R_1 \rightarrow R_1$ yields:

$$\begin{bmatrix} 1 & 0 & -1 & 4/3 & -10/3 & -5/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$-4/3R_3 + R_1 \rightarrow R_1$ yields:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & -4 & -11/3 \\ 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$1R_3 + R_2 \rightarrow R_2$ yields:

$$\begin{bmatrix} 1 & 0 & -1 & 0 & -4 & -11/3 \\ 0 & 1 & 2 & 0 & 7/2 & 5/2 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

which is our result. **q.e.d.**

Later in this class, we shall prove:

Fact 5.18 (uniqueness of the reduced row-echelon form). *Every matrix is equivalent to one and only one matrix in reduced row-echelon form.*

5.3 Linear Systems and Vector Spaces

Observation 5.19. *A linear system in s variables $x_1, x_2, \ldots, x_s$ with r equations can be written as a single equation*

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w}$$

using vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s, \mathbf{w} \in \mathbb{R}^r$.

Example 5.20. Consider the system

$$\left\| \begin{array}{rrrl} x & +3y & -z & = 9 \\ 2x & +y & +3z & = 8 \end{array} \right\|$$

Note that we can write this as one equation using (column) vectors:

$$x \begin{bmatrix} 1 \\ 2 \end{bmatrix} + y \begin{bmatrix} 3 \\ 1 \end{bmatrix} + z \begin{bmatrix} -1 \\ 3 \end{bmatrix} = \begin{bmatrix} 9 \\ 8 \end{bmatrix}.$$

Definition 5.21. A linear system is homogeneous if its right hand side consists of zero entries only. The corresponding vector equation would therefore be

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0}$$

Proposition 5.22 (the kernel is a subspace). *The set*

$$S := \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \in \mathbb{R}^s \;\middle|\; x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0} \right\}$$

of solutions to a homogeneous system of equations is a subspace of $\mathbb{R}^s$.

Proof. We verify the hypotheses of the Subspace Theorem. First observe that

$$0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_s = \mathbf{0}$$

whence

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \in S.$$

Now, we consider two solutions

$$\begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix}, \begin{bmatrix} y_1 \\ \vdots \\ y_s \end{bmatrix} \in S$$

Then, we have

$$\begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} + \begin{bmatrix} y_1 \\ \vdots \\ y_s \end{bmatrix} = \begin{bmatrix} x_1 + y_1 \\ \vdots \\ x_s + y_s \end{bmatrix} \in S$$

since

$$\begin{aligned} (x_1 + y_1)\mathbf{v}_1 + \cdots + (x_s + y_s)\mathbf{v}_s &= x_1\mathbf{v}_1 + y_1\mathbf{v}_1 + \cdots + x_s\mathbf{v}_s + y_s\mathbf{v}_s \\ &= (x_1\mathbf{v}_1 + \cdots + x_s\mathbf{v}_s) + (y_1\mathbf{v}_1 + \cdots + y_s\mathbf{v}_s) \\ &= \mathbf{0} \end{aligned}$$

Thus, S is closed with respect to addition.

Similarly, we find that for a solution

$$\begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \in S$$

we have

$$r \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} = \begin{bmatrix} rx_1 \\ \vdots \\ rx_s \end{bmatrix} \in S$$

since

$$\begin{aligned} (rx_1)\mathbf{v}_1 + \cdots + (rx_s)\mathbf{v}_s &= r(x_1\mathbf{v}_1 + \cdots + x_s\mathbf{v}_s) \\ &= r\mathbf{0} \\ &= \mathbf{0} \end{aligned}$$

Thus, the Subspace Theorem applies. **q.e.d.**

Definition 5.23. The vectors $\mathbf{v}_1, \ldots, \mathbf{v}_s$ are called linearly dependent if the linear system

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0}$$

has a non-trivial solution. If it has only the trivial solution, we call these vectors linearly independent.

Exercise 5.24. Determine whether the vectors

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

are linearly independent. (Remark: it is pure coincidence that the number of variables and the number of equation are equal in this example!)

Proposition 5.25 (the image/span is a subspace). *Given vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s \in \mathbb{R}^r$*, the set*

$$S := \{\mathbf{w} \in \mathbb{R}^r \mid x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w} \text{ has a solution}\}$$

of all right hand sides for which the corresponding linear system is solvable, is a subspace. (See exercise 1.8.18)

Proof. We verify the hypotheses of the Subspace Theorem. First, we have $\mathbf{0} \in S$ since a homogeneous linear system of equation always has a solution (e.g., the trivial solution).

Now assume $\mathbf{w}_x, \mathbf{w}_y \in S$, i.e., there there are scalars $x_1, \ldots, x_s \in \mathbb{R}$ such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w}_x$$

and scalars $y_1, \ldots, y_s \in \mathbb{R}$ such that

$$y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_s\mathbf{w}_s = \mathbf{w}_y.$$

We observe

$$\begin{aligned}\mathbf{w}_x + \mathbf{w}_y &= (x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s) + (y_1\mathbf{w}_1 + y_2\mathbf{w}_2 + \cdots + y_s\mathbf{w}_s) \\ &= (x_1 + y_1)\mathbf{v}_1 + (x_2 + y_2)\mathbf{v}_2 + \cdots + (x_s + y_s)\mathbf{v}_s\end{aligned}$$

Hence, $\mathbf{w}_x + \mathbf{w}_y \in S$.

Similarly, we show that S is closed with respect to scalar multiplication: Again assume $\mathbf{w} \in S$, i.e., there there are scalars $x_1, \ldots, x_s \in \mathbb{R}$ such that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w}.$$

Given a scalar r, we find

$$\begin{aligned}r\mathbf{w} &= r(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s) \\ &= (rx_1)\mathbf{v}_1 + (rx_2)\mathbf{v}_2 + \cdots + (rx_s)\mathbf{v}_s\end{aligned}$$

Thus, the Subspace Theorem applies. **q.e.d.**

Exercise 5.26 (finding the image). Determine all triples $(a, b, c) \in \mathbb{R}^3$ such that

$$\left\| \begin{array}{rrrl} x & +2y & -z & = a \\ 2x & -y & & = b \\ 3x & -4y & +z & = c \end{array} \right\|$$

has a solution.

Solution 5.27. We row-reduce:

$$\begin{bmatrix} 1 & 2 & -1 & a \\ 2 & -1 & 0 & b \\ 3 & -4 & 1 & c \end{bmatrix}$$

$-2R_1 + R_2 \to R_2$

$$\begin{bmatrix} 1 & 2 & -1 & a \\ 0 & -5 & 2 & -2a+b \\ 3 & -4 & 1 & c \end{bmatrix}$$

$-3R_1 + R_3 \to R_3$

$$\begin{bmatrix} 1 & 2 & -1 & a \\ 0 & -5 & 2 & -2a+b \\ 0 & -10 & 4 & -3a+c \end{bmatrix}$$

$-\frac{1}{5}R_2 \to R_2$

$$\begin{bmatrix} 1 & 2 & -1 & a \\ 0 & 1 & -\frac{2}{5} & \frac{2}{5}a - \frac{1}{5}b \\ 0 & -10 & 4 & -3a+c \end{bmatrix}$$

$10R_2 + R_3 \to R_3$

$$\begin{bmatrix} 1 & 2 & -1 & a \\ 0 & 1 & -\frac{2}{5} & \frac{2}{5}a - \frac{1}{5}b \\ 0 & 0 & 0 & a-2b+c \end{bmatrix}$$

$-2R_2 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 0 & -\frac{1}{5} & \frac{1}{5}a + \frac{2}{5}b \\ 0 & 1 & -\frac{2}{5} & \frac{2}{5}a - \frac{1}{5}b \\ 0 & 0 & 0 & a-2b+c \end{bmatrix}$$

Thus, the system is solvable if and only if

$$a - 2b + c = 0.$$

Remark 5.28. Our computation establishes an interesting identity:

$$\left\{ x\begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} + y\begin{bmatrix} 2 \\ -1 \\ -4 \end{bmatrix} + z\begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \,\middle|\, x, y, z \in \mathbb{R} \right\} = \left\{ \begin{bmatrix} a \\ b \\ c \end{bmatrix} \in \mathbb{R}^3 \,\middle|\, a - 2b + c = 0 \right\}$$

This is an identity of subspaces in $\mathbb{R}^3$. On the left hand side, we describe the subspace from the inside: the vectors

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

all lie within the subspace, and the subspace is <u>spanned</u> by these vectors. On the right hand side, we describe the subspace a the set of solutions to a homogeneous system of linear equations. Row reduction actually allows us to convert freely between the two points of view. Make sure that you understand this technique.

Example 5.29 (Interpolation). Find a quadratic polynomial ax^2+bx+c through:

$$(1,3), \quad (2,1), \quad (3,-3)$$

We know:

$$\begin{aligned} a1^2 + b1 + c &= 3 \\ a2^2 + b2 + c &= 1 \\ a3^2 + b3 + c &= -3 \end{aligned}$$

and this is a linear system, which we solve. This way, one can always find a polynomial taking some prescribed values at finitely many given points.

5.4 Inhomogeneous Systems

Theorem 5.30. *Fix $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s, \mathbf{w} \in \mathbb{R}^r$. Let $V \subseteq \mathbb{R}^s$ be the set of solutions to the homogeneous system*

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0}$$

i.e.,

$$V = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \in \mathbb{R}^s \,\middle|\, x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0} \right\}.$$

Then, the set

$$A := \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \in \mathbb{R}^s \,\middle|\, x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w} \right\}$$

of solutions to the inhomogeneous system is either empty or of the form

$$A = \{\mathbf{x}_0 + \mathbf{v} \mid \mathbf{v} \in V\}$$

where $\mathbf{x}_0$ is any given solution to the inhomogeneous system. Such a set is called an affine subspace of $\mathbb{R}^s$.

We will prove this later as a special case of a more general theorem.

6 Dimension Theory

6.1 Spans and Linear Independence in Abstract Vector Spaces

We have seen how we can rewrite systems of linear equations as a single equation using vectors. This motivates the following:

Definition 6.1. Let V be a vector space, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s \in V$ be a finite collection of vectors. An expression of the form

$$r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_s\mathbf{v}_s$$

is called a linear combination of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s$. Given another vector $\mathbf{w} \in V$, we call

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w}$$

a linear equation in the unknowns $x_1, x_2, \ldots, x_s$.

We call such a linear equation homogeneous if the right hand side $\mathbf{w}$ happens to be the zero vector $\mathbf{0}$.

Remark 6.2. You should think of linear equations as generalized systems of linear equations. In general, one such vector-equation will boil down to a messy linear system.

Theorem 6.3. *The set*

$$S := \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_s \end{bmatrix} \in \mathbb{R}^s \,\middle|\, x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0} \right\}$$

of solutions to a homogeneous linear equation is a subspace of $\mathbb{R}^s$*.*

Proof. The proof of Proposition 5.22 applies verbatim. **q.e.d.**

Definition 6.4. Let V be a vector space, and let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s \in V$ be a finite collection of vectors. Then the span of these vectors is the set

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\} := \{r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_s\mathbf{v}_s \mid r_1, r_2, \ldots, r_s \in \mathbb{R}\}.$$

This is the set of right hand sides $\mathbf{w} \in V$ for which the linear system

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w}$$

is solvable.

Theorem 6.5. *Given vectors* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s \in V$*, the set*

$$\operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_s\} = \{\mathbf{w} \in V \mid x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{w} \text{ has a solution}\}$$

of all right hand sides for which the corresponding linear equation is solvable, is a subspace of V*.*

Proof. The proof of Proposition 5.25 applies verbatim to the slightly more general situation here. **q.e.d.**

Definition 6.6. Let V be a vector space, and let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_s \in V$ be a finite collection of vectors. We call these vectors linearly independent if the homogeneous linear equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_s\mathbf{v}_s = \mathbf{0}$$

has no non-trivial solutions, i.e., its only solution is

$$x_1 = x_2 = \cdots = x_s = 0.$$

Example 6.7. We determine whether the polynomials

$$(t-1)^2, t^2, (t+1)^2, (t+2)^2 \in \mathbb{P}_2$$

are linearly independent. Thus, we have to determine whether the homogeneous linear equation

$$x_1(t-1)^2 + x_2t^2 + x_3(t+1)^2 + x_4(t+2)^2 = 0$$

has a non-trivial solution. We get:

$$\begin{aligned} x_1(t-1)^2 + x_2t^2 + x_3(t+1)^2 + x_4(t+2)^2 &= \\ x_1\big(t^2 - 2t + 1\big) + x_2\big(t^2\big) + x_3(t_2 + 2t + 1) + x_4(t_2 + 4t + 4) &= \\ (x_1 + x_2 + x_3 + x_4)t^2 + (-2x_1 + 2x_3 + 4x_4)t + (x_1 + x_3 + 4x_4) &= 0 \end{aligned}$$

Since a polynomial vanishes if and only if all its coefficients are 0, this equation is equivalent to the homogeneous linear system

$$\left\Vert \begin{array}{rrrrl} x_1 & +x_2 & +x_3 & +x_4 & = 0 \\ -2x_1 & & +2x_3 & +4x_4 & = 0 \\ x_1 & & +x_3 & +4x_4 & = 0 \end{array} \right\Vert$$

Since this is a system in 4 variables with 3 equations, we can see right away that there will be non-trivial solutions. Thus the four polynomials above are not linearly independent.

Example 6.8. Let us modify the example slightly. We ask whether the polynomials

$$(t-1)^2, t^2, (t+1)^2 \in \mathbb{P}_2$$

are linearly independent. Thus, we have to determine whether the homogeneous linear equation

$$x_1(t-1)^2 + x_2t^2 + x_3(t+1)^2 = 0$$

has a non-trivial solution. We get:

$$\begin{aligned} x_1(t-1)^2 + x_2t^2 + x_3(t+1)^2 &= \\ x_1\big(t^2 - 2t + 1\big) + x_2\big(t^2\big) + x_3(t_2 + 2t + 1) &= \\ (x_1 + x_2 + x_3)t^2 + (-2x_1 + 2x_3)t + (x_1 + x_3) &= 0 \end{aligned}$$

Since a polynomial vanishes if and only if all its coefficients are 0, this equation is equivalent to the homogeneous linear system

$$\left\|\begin{array}{rrrr} x_1 & +x_2 & +x_3 & =0 \\ -2x_1 & & +2x_3 & =0 \\ x_1 & & +x_3 & =0 \end{array}\right\|$$

Note how all this just amounts to omitting the fourth column of the linear system we got previously. Obviously, in some way yet to be explained, the fourth column of that system corresponds to the forth polynomial.

Anyway, no we have to row-reduce this system. We get:

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

$2R_1 + R_2 \to R_2$

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

$-1R_1 + R_3 \to R_3$

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

$\frac{1}{2}R_2 \to R_2$

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

$1R_2 + R_3 \to R_3$

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

$\frac{1}{2}R_3 \to R_3$

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

$-1R_2 + R_1 \to R_1$

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

$1R_3 + R_1 \to R_1$

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

$-2R_3 + R_2 \to R_2$

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

Since each column has a leading 1 except for the right hand side, we see that the system has only the trivial solution. Thus the three polynomials are linearly independent.

Example 6.9. Again, we modify the example. Now, we ask whether the polynomials

$$(t-1)^2, t^2, (t+1)^2 \in \mathbb{P}_2$$

span all of $\mathbb{P}_2$. This is to ask whether every polynomial

$$at^2 + bt + c$$

can be written as a linear combination of the given three polynomials. I.e., we ask whether the linear equation

$$x_1(t-1)^2 + x_2t^2 + x_3(t+1)^2 = at^2 + bt + c$$

can always be solved. Running through the same computation as in the previous example, we find that this equation is equivalent to the linear system system

$$\left\| \begin{matrix} x_1 & +x_2 & +x_3 & = a \\ -2x_1 & & +2x_3 & = b \\ x_1 & & +x_3 & = c \end{matrix} \right\|$$

which we can row-reduce:

$$\begin{bmatrix} 1 & 1 & 1 & 1a+0b+0c \\ -2 & 0 & 2 & 0a+1b+0c \\ 1 & 0 & 1 & 0a+0b+1c \end{bmatrix}$$

$2R_1 + R_2 \to R_2$

$$\begin{bmatrix} 1 & 1 & 1 & 1a+0b+0c \\ 0 & 2 & 4 & 2a+1b+0c \\ 1 & 0 & 1 & 0a+0b+1c \end{bmatrix}$$

$-1R_1 + R_3 \to R_3$

$$\begin{bmatrix} 1 & 1 & 1 & 1a+0b+0c \\ 0 & 2 & 4 & 2a+1b+0c \\ 0 & -1 & 0 & -1a+0b+1c \end{bmatrix}$$

$\frac{1}{2}R_2 \to R_2$

$$\begin{bmatrix} 1 & 1 & 1 & 1a+0b+0c \\ 0 & 1 & 2 & 1a+\frac{1}{2}b+0c \\ 0 & -1 & 0 & -1a+0b+1c \end{bmatrix}$$

$1R_2 + R_3 \to R_3$

$$\begin{bmatrix} 1 & 1 & 1 & 1a + 0b + 0c \\ 0 & 1 & 2 & 1a + \frac{1}{2}b + 0c \\ 0 & 0 & 2 & 0a + \frac{1}{2}b + 1c \end{bmatrix}$$

$\frac{1}{2}R_3 \to R_3$

$$\begin{bmatrix} 1 & 1 & 1 & 1a + 0b + 0c \\ 0 & 1 & 2 & 1a + \frac{1}{2}b + 0c \\ 0 & 0 & 1 & 0a + \frac{1}{4}b + \frac{1}{2}c \end{bmatrix}$$

$-1R_2 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 0 & -1 & 0a - \frac{1}{2}b + 0c \\ 0 & 1 & 2 & 1a + \frac{1}{2}b + 0c \\ 0 & 0 & 1 & 0a + \frac{1}{4}b + \frac{1}{2}c \end{bmatrix}$$

$1R_3 + R_1 \to R_1$

$$\begin{bmatrix} 1 & 0 & 0 & 0a - \frac{1}{4}b + \frac{1}{2}b \\ 0 & 1 & 2 & 1a + \frac{1}{2}b + 0c \\ 0 & 0 & 1 & 0a + \frac{1}{4}b + \frac{1}{2}c \end{bmatrix}$$

$-2R_3 + R_2 \to R_2$

$$\begin{bmatrix} 1 & 0 & 0 & 0a - \frac{1}{4}b + \frac{1}{2}c \\ 0 & 1 & 0 & 1a + 0b - 1c \\ 0 & 0 & 1 & 0a + \frac{1}{4}b + \frac{1}{2}c \end{bmatrix}$$

From the solution, we gather that we can always solve for x_1, x_2, x_3 given any values for a, b, c. Thus the three polynomials span $\mathbb{P}_2$. Note that they do not stand the slightest chance of spanning $\mathbb{P}_3$ since you will not be able to write t^3 as a linear combination of polynomials of degree 2.

6.2 Coordinates

Proposition and Definition 6.10 (The Evaluation Map) *Let V be a vector space and fix a finite sequence S of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$. Then, the evaluation map*

$$\begin{array}{rcl} \mathrm{ev}_S : \mathbb{F}^r & \longrightarrow & V \\ \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_r \end{bmatrix} & \mapsto & r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_r\mathbf{v}_r \end{array}$$

is linear.

Proof. We check additivity first:

$$\begin{aligned} \mathrm{ev}_S\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_r \end{bmatrix} + \begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{bmatrix}\right) &= \mathrm{ev}_S\left(\begin{bmatrix} r_1 + s_1 \\ r_2 + s_2 \\ \vdots \\ r_r + s_r \end{bmatrix}\right) \\ &= (r_1 + s_1)\mathbf{v}_1 + (r_2 + s_2)\mathbf{v}_2 + \cdots + (r_r + s_r)\mathbf{v}_r \\ &= (r_1\mathbf{v}_1 + s_1 + \mathbf{v}_1) + (r_2\mathbf{v}_2 + s_2\mathbf{v}_2) + \cdots + (r_r\mathbf{v}_r + s_r\mathbf{v}_r) \\ &= (r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_r\mathbf{v}_r) + (s_1\mathbf{v}_1 + s_2\mathbf{v}_2 + \cdots + s_r\mathbf{v}_r) \\ &= \mathrm{ev}_S\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_r \end{bmatrix}\right) + \mathrm{ev}_S\left(\begin{bmatrix} s_1 \\ s_2 \\ \vdots \\ s_r \end{bmatrix}\right) \end{aligned}$$

Now we check that ev_S is homogeneous:

$$\begin{aligned} \mathrm{ev}_S\left(s\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_r \end{bmatrix}\right) &= \mathrm{ev}_S\left(\begin{bmatrix} sr_1 \\ sr_2 \\ \vdots \\ sr_r \end{bmatrix}\right) \\ &= (sr_1)\mathbf{v}_1 + (sr_2)\mathbf{v}_2 + \cdots + (sr_r)\mathbf{v}_r \\ &= s(r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_r\mathbf{v}_r) \\ &= s\,\mathrm{ev}_S\left(\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_r \end{bmatrix}\right) \end{aligned}$$

This completes the proof. **q.e.d.**

Definition 6.11. Let M and N be two sets and let $f : M \to N$ be a map. The map f is called onto if every $n \in N$ has at least one preimage , i.e., an element $m \in M$ satisfying $f(m) = n$. The map f is called 1-1 if every $n \in N$ has at most one preimage.

A map $h : N \to M$ is called an inverse of f if

$$h(f(m)) = m \qquad \text{for all } m \in M$$

and

$$f(h(n)) = n \qquad \text{for all } n \in N.$$

Proposition 6.12. *A map $f : M \to N$ has an inverse if and only if f is onto and* 1-1*. If $h : N \to M$ is an inverse for f, then f is an inverse for h. If a map has an inverse, the inverse is onto and* 1-1*.*

Proof. First suppose that f is onto and 1-1. Consider $n \in N$. Since f is onto, n has at least one preimage, and since f is 1-1, it has at most one preimage. Hence, n has exactly one preimage $m \in M$. We define $h(n) := m$. Thus, h assigns to each n its unique preimage in M.

Now, we verify that $h : N \to M$ is an inverse for f. First consider $m \in M$. Then $h(f(m))$ is the unique preimage of $f(m)$. On the other hand, m is a preimage of $f(m)$. Hence $h(f(m)) = m$. Now, consider n. Then $h(n)$ is a preimage of n. But this says $f(h(n)) = n$. Thus, h is an inverse for f. This proves: If a map is onto and 1-1, it has an inverse.

Now assume that f has an inverse h. To finish the proof of the first claim, we have to see that f is onto and 1-1. So, consider $n \in N$. The map f is onto as we can find a preimage, namely $h(n)$. It remains to see, that f is 1-1. So suppose that m_1 and m_2 are both preimages of n. Then

$$m_1 = h(f(m_1)) = h(n) = h(f(m_2)) = m_2.$$

This concludes the proof of our first claim.

The second claim (if h is an inverse for f, then f is an inverse for h) is obvious: the two defining identities for being an inverse are symmetric with respect to changing the roles of f and h.

Finally, the last claim is an immediate consequence of the first two. **q.e.d.**

Theorem 6.13. *Let V be a vector space and fix a finite sequence S of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in V$. Then:*

1. *The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ span V if and only if the associated evaluation map $\mathrm{ev}_S : \mathbb{F}^r \to V$ is onto.*
2. *The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are linearly independent if and only if the associated evaluation map $\mathrm{ev}_S : \mathbb{F}^r \to V$ is* 1-1*.*

In particular, if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ is a basis of V, the associated evaluation map is onto and 1-1*.*

Proof. We start with claim 1. Let us consider the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r = \mathbf{w}.$$

By definition, the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ span V if this equation has a solution for every right hand side $\mathbf{w} \in V$. Note, however, that the above equation can be rewritten as

$$\mathrm{ev}_S\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}\right) = \mathbf{w}$$

which is solvable if and only if $\mathbf{w}$ has a preimage under the evaluation map. Thus it is solvable for every right hand side $\mathbf{w}$ if and only if the ev_S is onto.

To prove claim 2, we consider the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r = \mathbf{0}.$$

This equation can be rewritten as

$$\mathrm{ev}_S\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_r \end{bmatrix}\right) = \mathbf{0}$$

q.e.d.

Definition 6.14. A linear map that is onto and 1-1 is called an isomorphism.

Theorem 6.15. *The inverse of an isomorphism is linear. (Since the inverse is automatically onto and* 1-1 *by Proposition 6.12, we see that the inverse of an isomorphism is an isomorphism.)*

Proof. Let $\varphi : V \to W$ be an isomorphism; and let $\psi : W \to V$ be an inverse map for φ. First, we check that ψ is additive. For $\mathbf{w}_1, \mathbf{w}_2 \in W$, we have

$$\varphi(\psi(\mathbf{w}_1)) + \varphi(\psi(\mathbf{w}_2)) = \mathbf{w}_1 + \mathbf{w}_2$$

Thus, by linearity of φ:

$$\varphi(\psi(\mathbf{w}_1) + \psi(\mathbf{w}_2)) = \varphi(\psi(\mathbf{w}_1)) + \varphi(\psi(\mathbf{w}_2)) = \mathbf{w}_1 + \mathbf{w}_2.$$

Hence $\psi(\mathbf{w}_1) + \psi(\mathbf{w}_2)$ is a preimage of $\mathbf{w}_1 + \mathbf{w}_2$. Since such a preimage is unique and given by the inverse map, we find

$$\psi(\mathbf{w}_1) + \psi(\mathbf{w}_2) = \psi(\mathbf{w}_1 + \mathbf{w}_2)\,.$$

As for scaling, we consider $s \in \mathbb{F}$ and $\mathbf{w} \in W$. By linearity of φ, we find

$$\varphi(s\psi(\mathbf{w})) = s\varphi(\psi(\mathbf{w})) = s\mathbf{w}.$$

Thus, $s\psi(\mathbf{w})$ is the unique preimage of $s\mathbf{w}$, whence

$$s\psi(\mathbf{w}) = \psi(s\mathbf{w})\,.$$

q.e.d.

Definition 6.16. Let $\mathcal{V}$ be a basis of V, i.e., a finite sequence of linearly independent vectors $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_m$ that span V. Then the evaluation map $\mathrm{ev}_{\mathcal{V}} : \mathbb{F}^m \to V$ is linear by Proposition 6.10, and it is onto and 1-1 by Theorem 6.13. Thus, the evaluation map for a basis is an isomorphism. Its inverse isomorphism is called the coordinate map and denoted as follows:

$$\begin{array}{rcl} []_{\mathcal{V}} : V & \longrightarrow & \mathbb{F}^m \\ \mathbf{v} & \mapsto & [\mathbf{v}]_{\mathcal{V}} \end{array}$$

The coordinate vector $[\mathbf{v}]_{\mathcal{V}} \in \mathbb{F}^m$ associated to a vector $\mathbf{v} \in V$ is characterized by:

$$[\mathbf{v}]_{\mathcal{V}} = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix} \iff r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_m\mathbf{v}_m = \mathbf{v}.$$

Note that the coordinate map is linear by Theorem 6.15, i.e., we have for any $\mathbf{v}, \mathbf{w} \in V$ and any $r \in \mathbb{F}$:

$$\begin{array}{rcl} [\mathbf{v} + \mathbf{w}]_{\mathcal{V}} & = & [\mathbf{v}]_{\mathcal{V}} + [\mathbf{w}]_{\mathcal{V}} \\ [r\mathbf{v}]_{\mathcal{V}} & = & r[\mathbf{v}]_{\mathcal{V}} \end{array}$$

6.3 The Basis Selection Algorithm

Strong Expansion Lemma 6.17. *Suppose* $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i$ *are linearly independent. Then for any vector* $\mathbf{v}_{i+1}$*, we have*

$$\mathbf{v}_{i+1} \notin \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$$
if and only if
$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1} \text{ are linearly independent.}$$

Proof. We will actually prove the following equivalent statement:

$$\mathbf{v}_{i+1} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$$
if and only if
$$\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1} \text{ are } \underline{\text{not}} \text{ linearly independent.}$$

So assume first $\mathbf{v}_{i+1} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$. Then there are scalars $r_1, \ldots, r_i$ such that

$$r_1\mathbf{v}_1 + \cdots + r_i\mathbf{v}_i = \mathbf{v}_{i+1}.$$

In this case, we have

$$r_1\mathbf{v}_1 + \cdots + r_i\mathbf{v}_i - \mathbf{v}_{i+1} = \mathbf{0},$$

which shows that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1}$ are not linearly independent.

For the other implication, we assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i, \mathbf{v}_{i+1}$ are not linearly independent, i.e., that

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_i\mathbf{v}_i + x_{i+1}\mathbf{v}_{i+1} = \mathbf{0}$$

has a non-trivial solution. Note that in this solution x_{i+1} cannot be 0: if it was, we could drop the last summand and obtain a non trivial solution to

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_i\mathbf{v}_i = \mathbf{0}$$

which would imply that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i$ are not linearly independent contrary to our hypothesis.

Since $x_{i+1} \neq 0$, we can solve for $\mathbf{v}_{i+1}$:

$$\mathbf{v}_{i+1} = -\frac{x_1}{x_{i+1}}\mathbf{v}_1 - \frac{x_2}{x_{i+1}}\mathbf{v}_2 - \cdots - \frac{x_i}{x_{i+1}}\mathbf{v}_i$$

and this shows

$$\mathbf{v}_{i+1} \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_i\}$$

as predicted. **q.e.d.**

Algorithm for Finding a Basis within a Spanning Set 6.18. *Let* V *be a vector space. Suppose that*

$$S = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)$$

is a finite spanning set for V*.*

To select a basis from this spanning set proceed as follows: working from lower indices to higher indices, check for each $\mathbf{v}_j$ *whether it is in the span of its predecessors. If so, drop* $\mathbf{v}_j$*, if not, keep* $\mathbf{v}_j$*. The basis* $\mathcal{V}$ *consists of the vectors you keep.*

Here is a more formal description of the method. Put

$$\begin{aligned} V_0 &:= \{\mathbf{0}\} \\ V_1 &:= \operatorname{span}\{\mathbf{v}_1\} \\ V_2 &:= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\} \\ V_3 &:= \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_3\} \\ &\vdots \\ V_r &:= \operatorname{span}\{\mathbf{v}_1, \dots, \mathbf{v}_r\} \end{aligned}$$

Then

$$\mathcal{V} := \{\mathbf{v}_j \mid \mathbf{v}_j \notin V_{j-1}\}$$

is a basis for V*.*

This algorithm is biased in favor of low indices: if an initial segment of the sequence $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ *is a linearly independent set of vectors, it will be included as an initial segment in the basis.*

Proof of correctness. We have to argue that the output is a linearly independent spanning set for V.

Before we start, note how $V_0 = \{\mathbf{0}\}$ prevents the algorithm from (erroneously) including the zero-vector $\mathbf{0}$ in its selection.

As for spanning V, note that the algorithm drops exactly those vector $\mathbf{v}_j$ that are contained in the span of their preceding segment. Thus these vectors are redundant and do not contribute to the span. Therefore, the selection that the algorithm put out will be a spanning set for V.

As for linear independence of the output, we use the Expansion Lemma. As the algorithms walks through the spanning set, pondering whether it should keep vector $\mathbf{v}_j$ or throw it out, it has already compiled a list of preceding vectors that it decided to keep. Assume for a moment that the algorithm did not screw up earlier, i.e., assume that the list is has compiled so far is linearly independent. Then the vector $\mathbf{v}_j$ get added only if it is not contained in the span of the vectors $\mathbf{v}_i$ for $i < j$. Since the compiled list does not contain any higher indices, we see that if $\mathbf{v}_j$ is added it was not contained in the subspace spanned by the vectors on the list. Thus, by the Expansion Lemma 6.17, the extended list (with $\mathbf{v}_j$ included) stays linearly independent.

So, if the algorithm starts with a linearly independent compilation, it will keep it that way. Formally, the algorithm starts with an empty list. This is considered linearly independent. If you feel uneasy about that, think of the first vector that is added. Just observe that it has to be non-zero. Therefore, when the algorithm pick the first vector, it creates a list of one linearly independent vector: a set of one vector is linearly independent if and only if the vector is non-zero.

Finally the bias statement also follows from the Expansion Lemma 6.17. However, this time the other direction of the equivalence is used: Since the vectors in the initial segment are linearly independent, the later are not contained in the span of their predecessors. Thus, the algorithm will keep them. **q.e.d.**

Theorem 6.19. *Let V be a vector space. Suppose that*

$$S = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r)$$

is a finite spanning set for V. Then S contains a basis for V. (In fact, any maximal linearly independent subset of S is a basis for V.)

Proof. Let algorithm 6.18 select a basis from the spanning set. **q.e.d.**

Theorem 6.20. *Suppose V has a finite basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$. Then every set of linearly independent vectors $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ can be extended to a basis for V.*

Proof. Consider the sequence

$$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r, \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m.$$

This sequence is a spanning set for V since the $\mathbf{v}_i$ already span V. Now use algorithm 6.18 to select a basis from this list. Since the initial segment $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$ consists of linearly independent vectors, the algorithm will keep these vectors: recall that the Basis Selection Algorithm is biased towards lower indices. Hence, the output of the algorithm will be a basis extending the list $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_r$. **q.e.d.**

6.4 Basis Selection in Standard Vector Spaces

Proposition 6.21 (Basis Selection in $\mathbb{R}^m$). *Let $V \subseteq \mathbb{R}^m$ be a subspace spanned by the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r \in \mathbb{R}^m$. Then the Basis Selection Algorithm 6.18 boils down to a single row-reduction: Let A be the matrix formed by the columns:*

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix}$$

Put A into reduced row-echelon form A^*. Keep exactly those $\mathbf{v}_j$ for which the j-column of A^* has a leading 1.

Proof by example. We find a basis for the subspace $V \subseteq \mathbb{R}^4$ spanned by the vectors

$$\mathbf{v}_1 = \begin{bmatrix} 0 \\ 0 \\ 3 \\ 6 \end{bmatrix}, \quad \mathbf{v}_2 = \begin{bmatrix} 1 \\ 0 \\ 4 \\ 7 \end{bmatrix}, \quad \mathbf{v}_3 = \begin{bmatrix} 2 \\ 0 \\ 5 \\ 8 \end{bmatrix}, \quad \mathbf{v}_4 = \begin{bmatrix} -1 \\ 2 \\ 0 \\ 1 \end{bmatrix}, \quad \mathbf{v}_5 = \begin{bmatrix} 3 \\ 1 \\ 2 \\ 1 \end{bmatrix}, \quad \mathbf{v}_6 = \begin{bmatrix} 1 \\ 3 \\ -1 \\ -3 \end{bmatrix}.$$

We have

$$A = \begin{bmatrix} 0 & 1 & 2 & -1 & 3 & 1 \\ 0 & 0 & 0 & 2 & 1 & 3 \\ 3 & 4 & 5 & 0 & 2 & -1 \\ 6 & 7 & 8 & 1 & 1 & -3 \end{bmatrix}$$

The following sequence of row-operations reduces A:

$$\begin{aligned} & R_1 \leftrightarrow R_3 \\ & 1/3R_1 \to R_1 \\ & -6R_1 + R_4 \to R_4 \\ & R_2 \leftrightarrow R_3 \\ & 1R_2 + R_4 \to R_4 \\ & 1/2R_3 \to R_3 \\ & -4/3R_2 + R_1 \to R_1 \\ & -4/3R_3 + R_1 \to R_1 \\ & 1R_3 + R_2 \to R_2 \end{aligned}$$

We obtain the reduced row-echelon form:

$$A^* = \begin{bmatrix} 1 & 0 & -1 & 0 & -4 & -11/3 \\ 0 & 1 & 2 & 0 & 7/2 & 5/2 \\ 0 & 0 & 0 & 1 & 1/2 & 3/2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

So, the Proposition predicts that the Basis Selection Algorithm will choose $\mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_4$ as a basis for V.

To verify this prediction, we walk through the algorithm:

ponder about $\mathbf{v}_1$: We have to check whether $\mathbf{v}_1$ is the zero-vector. Well, we can read that off from A as well as from the reduced row-echelon form A^*. The leading 1 in the first column of A^* says that $\mathbf{v}_1$ is non-zero. Therefore the Basis Selection Algorithm will include $\mathbf{v}_1$ in the basis, as predicted.

ponder about $\mathbf{v}_2$: The Basis Selection Algorithm will keep $\mathbf{v}_2$ if and only if $\mathbf{v}_2 \notin \operatorname{span}\{\mathbf{v}_1\}$, i.e., if and only if the inhomogeneous linear equation

$$x_1\mathbf{v}_1 = \mathbf{v}_2$$

has no solution. Writing out the equation, we obtain the linear system

$$\left\| \begin{aligned} 0x_1 &= 1 \\ 0x_1 &= 0 \\ 3x_1 &= 4 \\ 6x_1 &= 7 \end{aligned} \right\|$$

Of course, we can eyeball easily that this is an inconsistent system, however, we can also deduce this from the reduced row-echelon form A^*: Apply the sequence of row-operations reducing A to the matrix

$$A_{1,2} := \begin{bmatrix} 0 & 1 \\ 0 & 0 \\ 3 & 4 \\ 6 & 7 \end{bmatrix}$$

which encodes the linear equation above. The sequence of row-operations will yield:

$$A^*_{1,2} := \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \end{bmatrix}$$

as the reduced row-echelon form for $A_{1,2}$. The leading 1 in the RHS column tells us that the system is inconsistent. Thus the leading 1 in the second column of A^* shows that the Basis Selection Algorithm will keep $\mathbf{v}_2$ on the list.

ponder about $\mathbf{v}_3$: Here is what the Basis Selection Algorithm does: Check whether

$$\mathbf{v}_3 \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

If "yes", throw it out, if "no" keep it.

To decide whether

$$\mathbf{v}_3 \in \operatorname{span}\{\mathbf{v}_1, \mathbf{v}_2\}$$

we have to see whether

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 = \mathbf{v}_3$$

has a solution. The matrix encoding this system is

$$A_{1,2,3} = \begin{bmatrix} 0 & 1 & 2 \\ 0 & 0 & 0 \\ 3 & 4 & 5 \\ 6 & 7 & 8 \end{bmatrix}$$

and the same sequence of row-operation that we already used will reduce it to:

$$A^*_{1,2,3} = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 2 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Now, the RHS column does not have a leading 1. Instead we can read off that

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

Thus, the fact that the third column of A^* does not have a leading 1 tells us that the algorithm will drop $\mathbf{v}_3$, as predicted.

ponder about $\mathbf{v}_4$: Now the question is whether

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 = \mathbf{v}_4$$

has a solution, and the first four columns of A encode this linear system:

$$A_{1,2,3,4} = \begin{bmatrix} 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 2 \\ 3 & 4 & 5 & 0 \\ 6 & 7 & 8 & 1 \end{bmatrix}$$

Again, we use the same sequence of row-operations and obtain the reduce row-echelon form

$$A^*_{1,2,3,4} = \begin{bmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

The leading 1 in the RHS column (corresponding to the fourth column of A^*) tells us that the system is inconsistent, whence the algorithm includes $\mathbf{v}_4$. A leading 1 forces inclusion.

ponder about $\mathbf{v}_5$: The algorithm needs to know whether

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + x_4\mathbf{v}_4 = \mathbf{v}_5$$

has a solution, and the matrix

$$A^*_{1,2,3,4,5} = \begin{bmatrix} 0 & 1 & 2 & -1 & 3 \\ 0 & 0 & 0 & 2 & 1 \\ 3 & 4 & 5 & 0 & 2 \\ 6 & 7 & 8 & 1 & 1 \end{bmatrix}$$

encodes the pertinent linear system. Our beloved sequence of row-operation works on this one just fine. We get: the reduced row-echelon form:

$$A^*_{1,2,3,4,5} = \begin{bmatrix} 1 & 0 & -1 & 0 & -4 \\ 0 & 1 & 2 & 0 & 7/2 \\ 0 & 0 & 0 & 1 & 1/2 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Now, the RHS column (that is the fifth column of A^*) does not contain a leading 1. Therefore, $\mathbf{v}_5$ is in the span of the first four vectors, and the algorithm will drop this redundant vector from the list.

ponder about $\mathbf{v}_6$: Finally, the matrix A itself encodes the linear equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + x_3\mathbf{v}_3 + x_4\mathbf{v}_4 + x_5\mathbf{v}_5 = \mathbf{v}_6$$

The sixth column of its reduced row-echelon form A^* corresponds to the RHS. It does not contain a leading 1. Therefore, the equation has a solution (which we can read off, if we are so inclined). This tells the algorithm that $\mathbf{v}_6$ is redundant and shall be dropped.

The example illustrates how we can use the same sequence of row-reductions to solve all the individual decision problems that the Basis Selection Algorithm requires us to tackle. That is the reason that we can combine all those computation in one big row-reduction. **q.e.d.**

It may be surprising, but now we are actually set to attack the following:

Theorem 6.22. *The reduced row-echelon form of any matrix is unique, i.e., there is only one possible result of row-reduction regardless of the sequence of elementary row-operations that one might choose.*

Proof by reusing the example. The key idea is to exploit the significance of the reduced row-echelon form for the Basis Selection Algorithm. Again let A^* be a reduced row-echelon form for A – we are assuming there might be more than one.

First note that the columns of A^* that have leading 1s correspond to the vectors that the Basis Selection Algorithm will select. Now that sequence is well-defined. Thus the output of the Basis Selection Algorithm can conversely be used to predict the columns with leading 1s in A^*. This settles the stair-case pattern of A^*.

In order to show that the remaining entries also are determined, we need to find interpretations of those numbers. The interpretation is: the numbers in the i-column of A^* are the coordinates of $\mathbf{v}_i$ relative to the basis $\mathcal{V}$ selected by the Basis Selection Algorithm. Thus, these numbers are uniquely determined by the sequence of column vectors forming A. **q.e.d.**

Another consequence of our understanding of how the Basis Selection Algorithm interacts with row-reduction is the following:

Theorem 6.23. *1. Every linearly independent set of vectors in $\mathbb{R}^m$ has at most m elements.*

2. Every spanning set of $\mathbb{R}^m$ has at least m elements.

3. Every basis of $\mathbb{R}^m$ has exactly m elements.

4. Every spanning set for $\mathbb{R}^m$ contains a basis.

5. Every linearly independent set of vectors in $\mathbb{R}^m$ can be extended to a basis.

Proof. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ be a list of column vectors in $\mathbb{R}^m$. We form the $m \times r$-matrix

$$A = \begin{bmatrix} \mathbf{v}_1 & \mathbf{v}_2 & \cdots & \mathbf{v}_r \end{bmatrix}$$

and we denote its reduced row-echelon form by A^*.

Assume $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are linearly independent. Then the Basis Selection Algorithm must keep all the vectors. Thus every column in A^* has a leading 1. Therefore, the matrix must have at least as many rows as it has columns, i.e, we have $m \geq r$. This proves the first claim.

Assume $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ span $\mathbb{R}^m$. Then every row in A^* has a leading 1. Hence A^* must have at least as many columns as it has rows. Therefore, $r \geq m$. This proves the second claim.

The third claim is now immediate: a basis is a set of linearly independent vectors, hence it has at most m vectors; on the other hand, a basis must span $\mathbb{R}^m$ and then it contains at least m vectors.

The last two claims just restate Theorems 6.19 and 6.20. **q.e.d.**

6.5 The Dimension of a Vector Space

Theorem and Definition 6.24. *Let V be a vector space. If V has a finite basis*

$$\mathcal{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$$

consisting of m vectors, then every other basis also has exactly m vectors. The number m of vectors in a basis for V is thus well-defined and called the dimension of V and denoted by $\dim(V)$. If V does not have a finite basis, we say that V is of infinite dimension: in this case, you can find finite independent sets of any size inside V.

Definition 6.25. A linear map that is onto and 1-1 is called an isomorphism.

Proposition 6.26. *Let $\varphi : V \longrightarrow W$ be linear.*

1. *If φ is onto, then it takes spanning sets for V to spanning sets for W; i.e., if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ spans V then $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ spans W.*

2. *If φ is 1-1, then it takes linearly independent sets in V to linearly independent sets in W; i.e., if $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ is a linearly independent set of vectors in V then $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ is a linearly independent set of vectors in W.*

Thus isomorphisms preserve the notions of spanning sets and linear independence.

In particular the coordinate map $[]_{\mathcal{V}} : V \longrightarrow \mathbb{R}^m$ and its inverse, the evaluation map, $\mathrm{ev}_{\mathcal{V}} : \mathbb{R}^m \longrightarrow V$ of a vector space with a finite basis $\mathcal{V}$ preserve the notions of spanning sets and linear independence.

Proof. We start with part 1. So assume $\varphi : V \rightarrow W$ is onto and that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ is a spanning set for V. We have to show that $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ is spanning set for W, i.e., we have to show that every vector $\mathbf{w} \in W$ can be written as a linear combination of the vectors $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$. So consider your favorite $\mathbf{w} \in W$. Since φ is onto, there is a vector $\mathbf{v} \in V$ with $\varphi(\mathbf{v}) = \mathbf{w}$; and since $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ spans V, there are scalars $r_1, r_2, \ldots, r_r \in \mathbb{F}$ such that

$$\mathbf{v} = r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_r\mathbf{v}_r.$$

Now, we remember that φ is linear. We get:

$$\begin{aligned}
\mathbf{w} &= \varphi(\mathbf{v}) \\
&= \varphi(r_1\mathbf{v}_1 + r_2\mathbf{v}_2 + \cdots + r_r\mathbf{v}_r) \\
&= \varphi(r_1\mathbf{v}_1) + \varphi(r_2\mathbf{v}_2) + \cdots + \varphi(r_r\mathbf{v}_r) \\
&= r_1\varphi(\mathbf{v}_1) + r_2\varphi(\mathbf{v}_2) + \cdots + r_r\varphi(\mathbf{v}_r)
\end{aligned}$$

This, however, represents $\mathbf{w}$ as a linear combination of the vectors $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ just as required.

Now for part 2. Here, we assume that φ is 1-1 and that the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ are linearly independent. We have to show that their images $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ are linearly independent. The crucial claim is the following:

Any solution $(x_1, x_2, \ldots, x_r)$ to the equation

$$x_1\varphi(\mathbf{v}_1) + x_2\varphi(\mathbf{v}_2) + \cdots + x_r\varphi(\mathbf{v}_r) = \mathbf{0}_W$$

is also a solution to the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r = \mathbf{0}_V.$$

Given this claim, we see immediately that if the later equation only has the trivial solution, then so has the former. Thus linear independence of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_r$ implies that the vectors $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_r)$ are linearly independent. To check the claim, we use that φ is 1-1: Let $(x_1, x_2, \ldots, x_r)$ be a solution to the equation

$$x_1\varphi(\mathbf{v}_1) + x_2\varphi(\mathbf{v}_2) + \cdots + x_r\varphi(\mathbf{v}_r) = \mathbf{0}_W.$$

Now consider the vector

$$\mathbf{v} := x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r \in V.$$

We know that

$$\begin{aligned} \varphi(\mathbf{v}) &= \varphi(x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r) \\ &= \varphi(x_1\mathbf{v}_1) + \varphi(x_2\mathbf{v}_2) + \cdots + \varphi(x_r\mathbf{v}_r) \\ &= x_1\varphi(\mathbf{v}_1) + x_2\varphi(\mathbf{v}_2) + \cdots + x_r\varphi(\mathbf{v}_r) \\ &= \mathbf{0}_W \end{aligned}$$

Since φ is 1-1, the zero-vector $\mathbf{0}_W$ has only one preimage: the zero-vector $\mathbf{0}_V$. Thus, we infer

$$\mathbf{0}_V = \mathbf{v} = x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r.$$

This proves that $(x_1, x_2, \ldots, x_r)$ also solves the equation

$$x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_r\mathbf{v}_r = \mathbf{0}_V$$

as promised. **q.e.d.**

Proof of Theorem 6.24. As in the theorem, we assume that V has a finite basis $\mathcal{V} = (\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m)$. The associated coordinate map

$$\begin{aligned} [\,]_{\mathcal{V}} : V &\longrightarrow \mathbb{R}^m \\ \mathbf{v} &\mapsto [\mathbf{v}]_{\mathcal{V}} \end{aligned}$$

is an isomorphism, i.e., it is linear, onto, and 1-1. Thus, $[\,]_{\mathcal{V}}$ takes any basis of V is taken to a basis of $\mathbb{R}^m$. However, bases in $\mathbb{R}^m$ all have exactly m elements. Hence the same holds for the bases in V. **q.e.d.**

Theorem 6.27. *Let V be a vector space of finite dimension m. Then the following hold:*

1. *V is isomorphic to $\mathbb{R}^m$. (There is essentially but one vector space for each dimension.)*

2. *Every basis of V has exactly m elements.*

3. *Every linearly independent set of vectors in V has at most m elements and can be extended to a basis.*

4. *Every spanning set of V has at least m elements and contains a basis.*

5. *Every subspace $S \subseteq V$ has finite dimension and $\dim(S) \leq m$.*

Proof. Fix a basis $\mathcal{V}$ for V. Then the coordinate map

$$[]_{\mathcal{V}} : V \longrightarrow \mathbb{R}^m$$

is an isomorphism. This proves the first claim. Except for the very last statement, the other claims now follow from their counterparts about $\mathbb{R}^m$ which have been established in Theorem 6.23.

As for the last claim, just observe that any linearly independent subset of S is a linearly independent subset of V and, therefore, can be extended to a basis. Thus m is an upper bound for the size of any linearly independent set in S. **q.e.d.**

Theorem 6.28. *Let $\varphi : V \longrightarrow W$ be linear.*

1. *If φ is onto, then $\dim(W) \leq \dim(V)$.*

2. *If φ is 1-1, then $\dim(W) \geq \dim(V)$.*

In particular, if φ is an isomorphism, we have $\dim(V) = \dim(W)$.

Proof. In the first claim, the inequality is only a restriction if V is of finite dimension. Thus, we may assume that V has a finite basis $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ where $m = \dim(V)$. Then, $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_m)$ is a spanning set for W since φ is onto. Thus $m \geq \dim(W)$ since the size of any spanning set for W is at least $\dim(W)$.

As for the second claim, we may assume that W has finite dimension (otherwise the claim is vacuous). Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ be a linearly independent set of vectors in V. Since φ is 1-1, the vectors $\varphi(\mathbf{v}_1), \varphi(\mathbf{v}_2), \ldots, \varphi(\mathbf{v}_m)$ form a linearly independent set in W. The size of such a set is bounded from above by $\dim(W)$. We infer $m \leq \dim(W)$. Thus, $\dim(W)$ is an upper bound for any linearly independent set in V. Hence the dimension of V cannot exceed the dimension of W. **q.e.d.**

Exercise 6.29. Let $V \subseteq \mathbb{P}_4$ be the set of all polynomials $p(t) = a_4t^4 + a_3t^3 + a_2t^2 + a_1t + a_0$ satisfying the following conditions simultaneously:

1. $p(-1) = 0$.

2. $p(3) = 0$.

3. $p'(1) = 0$.

Show that V is a subspace of $\mathbb{P}_4$ and determine the dimension of V.

7 Matrix Algebra

7.1 Matrix Multiplication

Example 7.1 (Input/Output tables for Bread). Here is a table showing two of my bread recipes:

	Wheat	Rye
wheat flour (oz)	24	16
rye flour (oz)	0	8
butter (tb)	1	1.5
milk (oz)	0	5
water (oz)	5	0
salt (tb)	0.5	0.5
suggar (tb)	1	2
yeast (tb)	1	1
eggs	3	3

And here is a table showing some prices (totally fictional):

	w. flour	r. flour	butter	milk	water	salt	sugar	yeast	egg
$	0.03	0.03	0.20	0.15	0.00	0.01	0.01	0.20	0.20

There is an obvious way to combine these data to obtain the prices for my breads: The price for a wheat bread is

$$0.03\times24+0.03\times0+0.20\times1+0.15\times0+0.00\times5+0.01\times0.5+0.01\times1+0.20\times1+0.20\times3$$

and the total price for a rye bread is

$$0.03\times16+0.03\times8+0.20\times1.5+0.15\times5+0.00\times0+0.01\times0.5+0.01\times2+0.20\times1+0.20\times3$$

We can display this in another table:

	Wheat	Rye
$	1.735	2.595

Definition 7.2 (Matrix Multiplication). Let A and B be two matrices. The product AB is defined whenever the number of columns in A matches the number of rows in B. In this case, the product has the same number of rows as the first factor (A) and the same number of columns as the second factor (B). Then entry of AB in row i and column j is:

$$\sum_{k=1}^{r} a_{i,k} b_{k,j}.$$

Here, r is the number of columns in A (and the number of rows in B).

Example 7.3 (Matrix Multiplication).

$$\begin{bmatrix} 2 & 3 \\ -2 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 2 & 0 & -1 & -1 \\ 3 & 4 & 1 & 1 \end{bmatrix} = \begin{bmatrix} * & * & * & * \\ * & * & (-2)\times(-1)+1\times 1 & * \\ * & * & * & * \end{bmatrix}$$

Theorem 7.4. *Let A and A' be $r \times s$ matrices, let B and B' be $s \times t$ matrices, and let C be an $t \times u$ matrix. Then*

1. $(AB)C = A(BC)$
2. $(A + A')B = AB + A'B$
3. $A(B + B') = AB + AB'$

I postpone the proof until we have some better tools. However, I shall give a reason why matrix multiplication ought to be associative.

Example 7.5 (Bread Revisited). Suppose someone commissions 3 wheat breads and 4 rye breads. There are two obvious ways of computing the total price of this order:

1. We can first compute the total ingredients for this order:

wheat flour (oz)	$24 \times 3 + 16 \times 4$	$= 136$
rye flour (oz)	$0 \times 3 + 8 \times 4$	$= 32$
butter (tb)	$1 \times 3 + 1.5 \times 4$	$= 9$
milk (oz)	$0 \times 3 + 5 \times 4$	$= 20$
water (oz)	$5 \times 3 + 0 \times 4$	$= 15$
salt (tb)	$0.5 \times 3 + 0.5 \times 4$	$= 3.5$
suggar (tb)	$1 \times 3 + 2 \times 4$	$= 11$
yeast (tb)	$1 \times 3 + 1 \times 4$	$= 7$
eggs	$3 \times 3 + 3 \times 4$	$= 21$

And now, we can compute the total price:

$$\begin{bmatrix} 0.03 & 0.03 & 0.20 & 0.15 & 0.00 & 0.01 & 0.01 & 0.20 & 0.20 \end{bmatrix} \begin{bmatrix} 136 \\ 32 \\ 9 \\ 20 \\ 15 \\ 3.5 \\ 11 \\ 7 \\ 21 \end{bmatrix} = \begin{bmatrix} 15.585 \end{bmatrix}$$

Note, that this amounts to computing the triple matrix product

$$\begin{bmatrix}0.03 & 0.03 & 0.20 & 0.15 & 0.00 & 0.01 & 0.01 & 0.20 & 0.20\end{bmatrix}\left(\begin{bmatrix}24 & 16\\ 0 & 8\\ 1 & 1.5\\ 0 & 5\\ 5 & 0\\ 0.5 & 0.5\\ 1 & 2\\ 1 & 1\\ 3 & 3\end{bmatrix}\begin{bmatrix}3\\ 4\end{bmatrix}\right)$$

2. We can use the precomputed table of prices:

	Wheat	Rye
$	1.537	2.397

Now, we get the total price by this matrix multiplication:

$$\begin{bmatrix}1.735 & 2.595\end{bmatrix}\begin{bmatrix}3\\ 4\end{bmatrix} = \begin{bmatrix}15.585\end{bmatrix}$$

And this amount to:

$$\left(\begin{bmatrix}0.03 & 0.03 & 0.20 & 0.15 & 0.00 & 0.01 & 0.01 & 0.20 & 0.20\end{bmatrix}\begin{bmatrix}24 & 16\\ 0 & 8\\ 1 & 1.5\\ 0 & 5\\ 5 & 0\\ 0.7 & 0.7\\ 1 & 2\\ 1 & 1\\ 2 & 2\end{bmatrix}\right)\begin{bmatrix}3\\ 4\end{bmatrix}$$

Since both ways ought to compute the total price, they better yield the same results. This is, why associativity should hold.

Remark 7.6. Don't be fooled by Theorem 7.4! Matrix multiplication is wicked.

1. Matrix multiplication is not commutative:

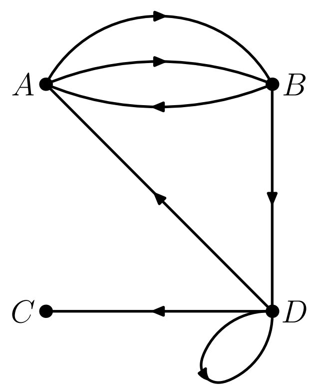
$$\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}1 & 1\\ 0 & 1\end{bmatrix} = \begin{bmatrix}0 & 1\\ 1 & 1\end{bmatrix}$$

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

2. To make things even worse, the product of two non-zero matrices can be the zero-matrix. Even the square of a non-zero matrix can vanish:

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

Example 7.7 (Directed Graphs). A directed graph consists of vertices and arrows connecting them. Here is an example with four vertices:



We can encode the information about this directed graph within a matrix as follows: We order the vertices. Then the entry $a_{i,j}$ represents the number of edges going from vertex i to vertex j. When A is the first, B is second, C is third, and D is the last vertex, we obtain the matrix

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

For instance, the first row says that A has exactly two outgoing edges, and both point towards B. The third row says that C has no outgoing edges at all. The Third column says that C has exactly one incoming edge, which issues from D.

Here is a remarkable fact: the power A^n computes the number of directed edge paths of length n. E.g., if you want to know how many routes there are from A to D of length exactly 100, you compute A^{100} and look at the entry in the first row and the fourth column.

Exercise 7.8. Compute A^4 and verify the above claim for some pairs of vertices.

7.2 The Matrix of a Linear Map

Observation 7.9. *For any $m \times n$-matrix A, the map*

$$\begin{aligned} \lambda_A : \mathbb{F}^n &\longrightarrow \mathbb{F}^m \\ \mathbf{x} &\longmapsto A\mathbf{x} \end{aligned}$$

is linear. In fact, given the definition of matrix multiplicatin, this map is no other than the linear map from Proposition 4.7.

Consequently, Proposition 4.10 implies that every linear map from $\mathbb{F}^n$ to $\mathbb{F}^m$ is of the form λ_A for some matrix A. **q.e.d.**

Theorem 7.10. *Let A be an $m \times n$-matrix with reduced row echelon form A^*. Then the following hold:*

1. *The linear map $\lambda_A : \mathbb{F}^n$ is onto if and only if the columns of A span $\mathbb{F}^m$, i.e., each row in the reduced row echelon form A^* has a leading 1, i.e.,*

2. *The linear map $\lambda_A : \mathbb{F}^n$ is 1-1 if and only if the columns of A are linearly independent, i.e., each column in the reduced row echelon form A^* has a leading 1.*

Proof. Let $\mathbf{a}_1, \ldots, \mathbf{a}_n \in \mathbb{F}^m$ be the columns of A. Note that $\mathbf{a}_i = A\mathbf{e}_i$ where $\mathbf{e}_i$ is the i^{th} standard basis vector. Then the map λ_A takes the standard basis to a spanning set for its image. Hence λ_A is onto if and only if the columns of A span $\mathbb{F}^m$.

Similarly,

$$\ker(\lambda_A) = \left\{ \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \,\middle|\, x_1\mathbf{a}_1 + \cdots + x_n\mathbf{a}_n = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \right\}$$

which is trivial if and only if the columns $\mathbf{a}_i$ are linearly independent. **q.e.d.**

7.3 Matrix Multiplication and Composition of Linear Maps

Compositions of linear maps are linear:

Proposition 7.11. *Let $\mathbb{F}$ be a field. Let V, W, and U be $\mathbb{F}$-vector spaces. Let $\varphi : V \to W$ and $\psi : W \to U$ be linear maps. Then the composition*

$$\begin{aligned} \psi \circ \varphi : V &\longrightarrow U \\ \mathbf{v} &\longmapsto \psi(\varphi(\mathbf{v})) \end{aligned}$$

is linear.

Proof. We show compatibility with addition. for $\mathbf{v}_1, \mathbf{v}_2 \in V$, we have:

$$\begin{aligned} (\psi \circ \varphi)(\mathbf{v}_1 \oplus_V \mathbf{v}_2) &= \psi(\varphi(\mathbf{v}_1 \oplus_V \mathbf{v}_2)) && \text{definition of composition} \\ &= \psi(\varphi(\mathbf{v}_1) \oplus_W \varphi(\mathbf{v}_2)) && \varphi \text{ is linear} \\ &= \psi(\varphi(\mathbf{v}_1)) \oplus_U \psi(\varphi(\mathbf{v}_2)) && \psi \text{ is linear} \\ &= (\psi \circ \varphi)(\mathbf{v}_1) \oplus_U (\psi \circ \varphi)(\mathbf{v}_2) && \text{definition of composition} \end{aligned}$$

It remains to argue that $\psi \circ \varphi$ is compatible with scalar multiplication. For $\mathbf{v} \in V$ and $r \in \mathbb{F}$, we have

$$\begin{aligned}
(\psi \circ \varphi)(r \odot_V \mathbf{v}) &= \psi(\varphi(r \odot_V \mathbf{v})) && \text{definition of composition} \\
&= \psi(r \odot_W \varphi(\mathbf{v})) && \varphi \text{ is linear} \\
&= r \odot_U \psi(\varphi(\mathbf{v})) && \psi \text{ is linear} \\
&= r \odot_U (\psi \circ \varphi)(\mathbf{v}) && \text{definition of composition}
\end{aligned}$$

q.e.d.

Theorem 7.12. *We have*

$$\lambda_A \circ \lambda_B = \lambda_{AB}.$$

Proof. Exercise: you have to show that

$$(\lambda_A \circ \lambda_B)(\mathbf{x}) = (AB)\mathbf{x} = A(B\mathbf{x}) = \lambda_{AB}(\mathbf{x})$$

for each column vector $\mathbf{x}$. This is just a big mess with many indices: for concreteness let $A = (a_{ij})$ be an $r \times s$-matrix, let $B = (b_{jk})$ be an $s \times t$-matrix, and let $\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_t \end{bmatrix}$.

We compute the left hand:

$$(AB)\mathbf{x} = \begin{bmatrix} \sum_{k=1}^{t} \left(\sum_{j=1}^{s} a_{1j} b_{jk} \right) x_k \\ \vdots \\ \sum_{k=1}^{t} \left(\sum_{j=r}^{s} a_{1j} b_{jk} \right) x_k \end{bmatrix}$$

and we compute the right hand:

$$A(B\mathbf{x}) = \begin{bmatrix} \sum_{j=1}^{s} a_{1j} \left(\sum_{k=1}^{t} b_{jk} x_k \right) \\ \vdots \\ \sum_{j=1}^{s} a_{rj} \left(\sum_{k=1}^{t} b_{jk} x_k \right) \end{bmatrix}$$

Those are visibly equal. **q.e.d.**

Corollary 7.13. *Matrix multiplication is associative.*

Proof. Matrix multiplication corresponds to composition of functions. Composition of functions is associative. **q.e.d.**

7.4 Matrices for Abstract Linear Maps

Let V and W be two vector spaces with bases $\mathcal{V} = (\mathbf{v}_1, \ldots, \mathbf{v}_n)$ and $\mathcal{W} = (\mathbf{w}_1, \ldots, \mathbf{w}_n)$, and let $\varphi : V \to W$ be a linear map. The composition

$$\mathbb{F}^n \xrightarrow{\mathrm{ev}_{\mathcal{V}}} V \xrightarrow{\varphi} W \xrightarrow{[]_{\mathcal{W}}} \mathbb{F}^m$$

is linear and hence given as matrix multiplication λ_A for some matrix A. This matrix is the coordinate matrix of φ relative to the basis $\mathcal{V}$ and $\mathcal{W}$. Then, we have the following commutative diagram of vector spaces and linear maps:

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ {\scriptstyle []_{\mathcal{V}}}\big\downarrow\big\uparrow{\scriptstyle \mathrm{ev}_{\mathcal{V}}} & & {\scriptstyle []_{\mathcal{W}}}\big\downarrow\big\uparrow{\scriptstyle \mathrm{ev}_{\mathcal{W}}} \\ \mathbb{F}^n & \xrightarrow[\lambda_A]{} & \mathbb{F}^m \end{array}$$

Proposition 7.14. *The columns of the coordinate matrix are* $[\varphi(e_i)]_{\mathcal{W}}$.

Proof. Exercise. **q.e.d.**

7.5 Realizing Row Operations

We describe some basic types of matrices and investigate the effects of multiplying by these matrices on the left.

Definition 7.15. We call a square matrix X a row-op matrix if left-multiplication by X realizes a row-operation.

This section is entirely devoted to establishing the following:

Proposition 7.16. *There are enough row-op matrices: every row operation can be realized via left-multiplication by a matrix (which then will be a row-op matrix).*

Since there are three types of row-operations, we will have to describe three different types of row-op matrices.

7.5.1 Permutation Matrices

Definition 7.17. A square matrix is called a permutation matrix if each row and each column has exactly one 1 and all other entries are 0.

Example 7.18. Here is a complete list of all 2×2 permutation matrices:

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

Example 7.19. The complete list of 3×3 permutation matrices is here:

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

Exercise 7.20. Show that, for every r, there are exactly

$$r! := r(r-1)(r-2)\cdots 1$$

permutation matrices of shape $r \times r$.

Example 7.21. Use your fingers to verify:

$$\underset{P}{\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}} \underset{A}{\begin{bmatrix} 2 & 4 & 0 & 1 \\ 4 & -6 & 1 & 2 \\ 3 & 1 & -1 & 0 \end{bmatrix}} = \underset{PA}{\begin{bmatrix} 4 & -6 & 1 & 2 \\ 3 & 1 & -1 & 0 \\ 2 & 4 & 0 & 1 \end{bmatrix}}.$$

Let us call the 3×3 permutation matrix on the left P. Now observe how the 1 in the first row of P picks the second row of the other factor in this matrix product (let us call that matrix A). The 0-entries in the first row of P are crucial in that they ensure that the first and third entry of each column in A gets ignored when computing the first row of the product.

Similarly, the second row of P is carefully designed so that the second row of the product will match the third row of A.

Finally the last row of P has its 1 in the first slot. Hence it copies the first row of A into the third row of the product.

We summarize the gist of this example computation:

Observation 7.22. *Left multiplication by a permutation matrix rearranged the rows of the other factor. Any rearrangement can be realized by an appropriately chosen permutation matrix. In particular, the row operation of swapping two rows can be realized via left-multiplication by a permutation matrix.*

Exercise 7.23. Show that the set of all $r \times r$ permutation matrices is a group with respect to the binary operation of matrix multiplication. Hint: Since matrix multiplication is associative in general (not only for permutation matrices), you do not need to check this axiom.

Remark 7.24. For $r \geq 3$ the group of all $r \times r$ permutation matrices is not abelian. We illustrate this for $r = 3$:

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

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

7.5.2 Elementary Matrices

Definition 7.25. A square matrix is called an <u>elementary matrix</u> if all diagonal entries are 1 and at most one off-diagonal entry is non-zero. Note that the identity matrices are elementary.

Example 7.26. Here is are some elementary matrix:

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

We want to understand what happens when we hit a matrix with an elementary matrix from the left. An example will enlighten us:

Example 7.27. Again, use your fingers and check:

$$\underset{E}{\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{bmatrix}} \underset{A}{\begin{bmatrix} 2 & 4 & 0 & 1 \\ 4 & -6 & 1 & 2 \\ 3 & 1 & -1 & 0 \end{bmatrix}} = \underset{EA}{\begin{bmatrix} 2 & 4 & 0 & 1 \\ -2 & -8 & 3 & 2 \\ 3 & 1 & -1 & 0 \end{bmatrix}}$$

Note how the first and last row of E just issue copy operation: they ensure that the first and the last row of the product EA match their counterparts in A exactly.

The second row of E is a little more tricky. Its 1-entry on the diagonal wants to just copy the second row from A, however the -2 in the third column interferes and we end up with a second row in the product EA where twice the third row of A got subtracted from its second row.

Take all effects and you see that this is just the row operation $R_2 - 2R_3 \rightarrow R_2$.

We generalize:

Observation 7.28. *The row-operation $R_i + CR_j \rightarrow R_i$ is realized via left-multiplication by the elementary matrix that has C in row i and column j.*

You should convince yourself by doing some more examples. Use your fingers!

7.5.3 Diagonal Matrices

Definition 7.29. A square matrix is called a diagonal matrix if all its off-diagonal entries are 0.

Example 7.30. Here are some 3×3 diagonal matrices:

$$\begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}, \begin{bmatrix} 3 & 0 & 0 \\ 0 & -\frac{2}{5} & 0 \\ 0 & 0 & -8 \end{bmatrix}.$$

Example 7.31. Time to use your fingers. Please check:

$$\underset{D}{\begin{bmatrix} 2 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}} \underset{A}{\begin{bmatrix} 2 & 4 & 0 & 1 \\ 4 & -6 & 1 & 2 \\ 3 & 1 & -1 & 0 \end{bmatrix}} = \underset{DA}{\begin{bmatrix} 4 & 8 & 0 & 2 \\ 2 & -3 & \frac{1}{2} & 1 \\ 3 & 1 & -1 & 0 \end{bmatrix}}$$

As you can see, the effect is rather predictable: every row of A gets multiplied by the corresponding diagonal entry from D.

Observation 7.32. *Left-multiplication by a diagonal matrix rescales every row by the corresponding diagonal entry. In particular, the row operation*

$$CR_i \rightarrow R_i$$

is realized by the diagonal matrix whose entry in row i and column i is C and whose other diagonal entries are 1.

7.6 Inverting Square Matrices

Definition 7.33. Let A be an $m \times m$ square matrix. We say that an $m \times m$ matrix L is a left-inverse for A if $LA = \mathbb{I}_m$. Similarly, we say that R is a right-inverse if $AR = \mathbb{I}_m$.

Observation 7.34. *Assume that L is a left-inverse for A and that R is a right-inverse. Then*

$$L = L\mathbb{I} = L(AR) = (LA)R = \mathbb{I}R = R.$$

In particular, if a matrix has left- and right-inverses, all those inverses coincide and the matrix has a unique (two-sided) inverse. **q.e.d.**

Algorithm 7.35. *To find the inverse of a $m \times m$ square matrix A, do the following*

1. *Form the matrix $[A|\mathbb{I}]$.*
2. *Apply row-operations to put the composite matrix into reduced row-echelon form.*
3. *If the reduced row echelon form looks like $[\mathbb{I}|B]$ then B is an inverse for A. Otherwise, A has neither a right- nor a left-inverse.*

Proof. First, assume that the algorithm terminates successfully, i.e., the reduced row echelon form has an identity matrix in the left block.

We begin by showing that in this case, the matrix B is a left-inverse for A. Let $X_1, X_2, \ldots, X_r$ be the row-op matrices realizing the sequence row-ops used to reduce A to $\mathbb{I}$. This means:

$$\mathbb{I} = X_r \cdots X_2 X_1 A.$$

Since B is obtained from applying these row-ops to $\mathbb{I}$, we find

$$B = X_r \cdots X_2 X_1 \mathbb{I} = X_r \cdots X_2 X_1.$$

Thus:

$$\mathbb{I} = BA$$

which implies that B is a left-inverse.

Now, we shall see that B is also a right-inverse for A. To see this, we interpret the i^{th} columnt $\mathbf{b}_i$ of B: delete all other columns from the right square block and apply the sequence of row-operations $X_1, \ldots, X_r$ to the matrix $[A|\mathbf{e}_i]$. The reduced row echelon form is $[\mathbb{I}|\mathbf{b}_i]$. Hence $\mathbf{b}$ is the unique solution to the equation

$$A\mathbf{x} = \mathbf{e}_i,$$

i.e., we have

$$A\mathbf{b}_i = \mathbf{e}_i.$$

Combining these identities for all columns, yields

$$AB = A[\mathbf{b}_1 \cdots \mathbf{b}_m] = [\mathbf{e}_1 \cdots \mathbf{e}_m] = \mathbb{I}.$$

q.e.d.

Theorem 7.36. *Let A be an $m \times m$ square matrix. Then the following are equivalent:*

1. *The linear map $\lambda_A : \mathbf{x} \mapsto A\mathbf{x}$ is an isomorphism.*
2. *The reduced row echelon form of A is the identity matrix.*
3. *A has a right-inverse.*
4. *A has a left-inverse.*

If any of these equivalent conditions is satisfied, the Algorithm 7.35 terminates succesfully and finds the inverse (which is unique by Observation 7.34).

Proof. We shall prove that all conditions are equivalent to the first. For the second, we have already seen this.

Also, the first conditions implies the last two: if λ_A is an isomorphism, then λ_A has an inverse, and this inverse is a linea map. The matrix B describing the inverse map will be a left- and a right-inverse to A.

Now assume that A has a right-inverse R. Then, the composition of functions $\lambda_{\mathbb{I}} = \lambda_{AR} = \lambda_A \circ \lambda_R$ is onto, which implies that λ_A is onto and therefore an isomorphism (see Theorem 7.10).

Finally, assume that A has a left-inverse L. Then, the composition of functions $\lambda_{\mathbb{I}} = \lambda_{LA} = \lambda_L \circ \lambda_A$ is 1-1, which implies that λ_A is 1-1 and therefore an isomorphism (see Theorem 7.10). **q.e.d.**

7.7 Rank and Nullity

Definition 7.37. Let A be an $m \times n$-matrix. The column-span of A is the subspace $\text{col-span}(A) \leq \mathbb{F}^m$ spanned by the columns of A. The row-span of A is the subspace $\text{row-span}(A) \leq \mathbb{F}^n$ spanned by the rows of A.

Observation 7.38. *Elementary row-operations do not change the row-span of a matrix. Hence, every matrix has the same row-span as its reduced row echelon form. In particular, the dimension of the row-span of any matrix is the number of leading* 1*s in its reduced row-echelon form.* **q.e.d.**

Observation 7.39. *The basis-selection algorithm will select a basis for the column-span of a matrix. Hence the dimension of the column-span of any matrix is the number of leading* 1*s in its reduced row-echelon form.* **q.e.d.**

This has a surprising consequence:

Corollary 7.40. *The row-span and the column-span of any matrix have the same dimension.* **q.e.d.**

Definition 7.41. Let A be an $m \times n$-matrix. Recall that A defines a linear map

$$\begin{array}{rcl} \lambda_A : \mathbb{F}^n & \longrightarrow & \mathbb{F}^m \\ \mathbf{x} & \mapsto & A\mathbf{x} \end{array}$$

1. The rank of A is defined as the dimension of the image of λ_A:

$$\text{rk}(A) := \dim(\text{im}(\lambda_A))$$

2. The nullity of A is defined as the dimension of its null-space, that is the kernel of λ_A:

$$\text{null}(A) := \dim(\ker(\lambda_A))$$

Observation 7.42. *The rank of a matrix A is the dimension of its column-span, since the column-span is the image of the linear map λ_A.* **q.e.d.**

Observation 7.43. *The nullity of an $m \times n$-matrix A is the number of free variable in the homogeneous system of equations* $A\mathbf{x} = \mathbf{0}$. *Since the rank is the number of dependent variables, we find that the sum of rank and nullity is the total number of variables (number of columns):* $\text{rk}(A) + \text{null}(A) = n$. **q.e.d.**

This extends to abstract linear maps:

Theorem 7.44. *Let $\varphi : V \to W$ be a linear map and assume that V and W have finite dimension. Then,* $\dim(V) = \dim(\text{im}(\varphi)) + \dim(\ker(\varphi))$.

Proof. Choosing bases $\mathcal{V}$ and $\mathcal{W}$ for V and W, we can represent φ by a matrix A:

$$\begin{array}{ccc} V & \xrightarrow{\varphi} & W \\ {\scriptstyle []_{\mathcal{V}}}\big\downarrow\big\uparrow{\scriptstyle \mathrm{ev}_{\mathcal{V}}} & & {\scriptstyle []_{\mathcal{W}}}\big\downarrow\big\uparrow{\scriptstyle \mathrm{ev}_{\mathcal{W}}} \\ \mathbb{F}^n & \xrightarrow[\lambda_A]{} & \mathbb{F}^m \end{array}$$

Now, the image of φ is isomorphic to the image of λ_A and the kernel of φ is isomorphic to the kernel of λ_A. Thus, the theorem follows from Observation 7.43. **q.e.d.**

8 The Determinant

8.1 Defintion of the Determinant

Definition 8.1. Let A be an $r \times r$ square matrix. The determinant $\det(A)$ is defined recursively as follows:

$r = 1$: We put

$$\det\big(a\big) := a$$

$r > 1$: We put

$$\det\begin{pmatrix} a_{1,1} & \cdots & a_{1,r} \\ \vdots & \ddots & \vdots \\ a_{r,1} & \cdots & a_{r,r} \end{pmatrix} := \sum_{i=1}^{r} (-1)^{i+1} a_{i,1} \det(A_{i,1})$$
$$= a_{1,1}\det(A_{1,1}) - a_{2,1}\det(A_{2,1}) + \cdots \pm a_{r,1}\det(A_{r,1})$$

where $A_{i,j}$ is the minor of A obtained by deleting the i^{th} row and the j^{th} column.

Note how the problem of computing a 3×3 determinant is reduced to 2×2 determinants (albeit three of them).

Example 8.2. For small matrices, it is easy to spell this out:

$$\det\begin{pmatrix} a & b \\ c & d \end{pmatrix} = a\det(d) - c\det(b) = ad - cb$$

$$\det\begin{pmatrix} a & b & c \\ d & e & f \\ g & h & i \end{pmatrix} = a\det\begin{pmatrix} e & f \\ h & i \end{pmatrix} - d\det\begin{pmatrix} b & c \\ h & i \end{pmatrix} + g\det\begin{pmatrix} b & c \\ e & f \end{pmatrix}$$
$$= a(ei - hf) - d(bi - hc) + g(bf - ec =$$
$$= aei - ahf + dhc - dbi + gbf - gec$$

Observe that each summand has r factors and the factors are taken from different rows and columns. (For those who play chess: each the summands correspond to the non-threatening rook configurations on an $r \times r$ board.)

That half of the summands occur with a negative sign is some black magic. (If you make all the signs positive, you obtain the so called permutant, which is by and large meaningless and for which no efficient way of computing it is known.)

8.2 Computing Determinants via Row-Reduction

We shall study how the determinant of a matrix A changes when we subject A to elementary row operations.

Exercise 8.3. The determinant is linear in each row. That means that if all but one rows of a matrix are fixes, the determinant is a linear function in the remaining row. So, let $R_1, \ldots, R_r, R_i'$ be row vectors and let r and s be scalars, then:

$$\det\begin{pmatrix} R_1 \\ \vdots \\ rR_i + sR_i' \\ \vdots \\ R_r \end{pmatrix} = r\det\begin{pmatrix} R_1 \\ \vdots \\ R_i \\ \vdots \\ R_r \end{pmatrix} + s\det\begin{pmatrix} R_1 \\ \vdots \\ R_i' \\ \vdots \\ R_r \end{pmatrix}$$

Lemma 8.4. *Let A be an $r \times r$ matrix.*

1. *Swapping two rows in A changes the sign of the determinant.*
2. *Multiplying a row of A by the scalar r multiplies the determinant by r.*
3. *Adding a multiple of a row in A to another row does not affect the determinant.*

Proof. Let A be an $r \times r$-matrix and let A' be obtained from A by the row-swap $R_i \leftrightarrow R_j$. Then

$$\det(A') = \sum_{k=1}^{r} (-1)^{1+k} a'_{k,1} \det\left(A'_{k,1}\right)$$

q.e.d.

Corollary 8.5. *Let X be an $r \times r$-matrix describing a row-operation. Then, for any $r \times r$-matrix A, we have* $\det(XA) = \det(X)\det(A)$.

Proof. A row-op-matrix is obtained by applying the row-op to the identity matrix, which has determinant 1. Hence, Lemma 8.4 implies:

- If X describes a row-swap, then $\det(X) = -1$.
- If X rescales a row by $r \neq 0$, then $\det(X) = r$.
- If X adds a multiple of a row to another row, then $\det(X) = 1$.

Now, we apply Lemma 8.4 to the matrix A and obtain:

- If X describes a row-swap, then $\det(XA) = -\det(A) = \det(X)\det(A)$.
- If X rescales a row by $r \neq 0$, then $\det(XA) = r\det(A) = \det(X)\det(A)$.
- If X adds a multiple of a row to another row, then $\det(X) = \det(A) = \det(X)\det(A)$.

This proves the claim in all three cases. **q.e.d.**

Corollary 8.6. *Let A be an $r \times r$-matrix and let $X_1, X_2, \ldots, X_r$ be a sequence of row-op-matrices row-reducing A so that $A^* := X_r \cdots X_2 X_1 A$ is in reduced row-echelon form. Then:*

1. *If A is not invertible, then $\det(A) = 0$.*
2. *If A is invertible, then $\det(A) = \frac{1}{\det(X_1)\det(X_2)\cdots\det(X_r)} \neq 0$.*

Proof. Corollary 8.5 implies:

$$\begin{aligned}\det(A^*) &= \det(X_r)\det(X_{r-1}\cdots X_1 A)\\ &= \det(X_r)\det(X_{r-1})\det(X_{r-2}\cdots X_1 A)\\ &= \cdots\\ &= \det(X_r)\det(X_{r-1})\cdots\det(X_1)\det(A)\end{aligned}$$

Now the claim follows from the observation that the reduced row echelon form A^* is the identity matrix if and only if A is invertible (in this case, A^* has determinant 1). If A is not invertible, the reduced row echelon form has a row of 0s and its determinant vanishes. **q.e.d.**

8.3 Multiplicativity of the Determinant

Theorem 8.7. *Let A and B be two $r \times r$ matrices. Then* $\det(AB) = \det(A)\det(B)$.

Proof. If either A or B are not invertible, then the product is not invertible and both sides of the identity vanish.

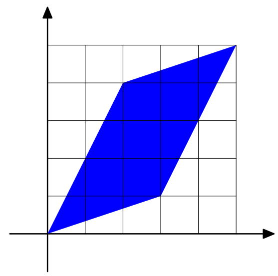
So now assume that A and B are both invertible. Let $X_1, X_2, \ldots, X_r$ be a sequence of row-op-matrices row-reducing A and let $Y_1, Y_2, \ldots, Y_s$ be a sequence of row-op-matrices row-reducing B. Note that the sequence $X_1, X_2, \ldots, X_r$ reduces AB to $X_r \cdots X_2, X_1 AB = B$. Hence the sequence $X_1, X_2, \ldots, X_r, Y_1, Y_2, \ldots, Y_s$ row-reduces AB and we obtain from Corollary 8.6

$$\begin{aligned}\det(AB) &= \frac{1}{\det(X_1)\cdots\det(X_r)\det(Y_1)\cdots\det(Y_s)}\\ &= \frac{1}{\det(X_1)\cdots\det(X_r)}\frac{1}{\det(Y_1)\cdots\det(Y_s)}\\ &= \det(A)\det(B)\end{aligned}$$

as claimed. **q.e.d.**

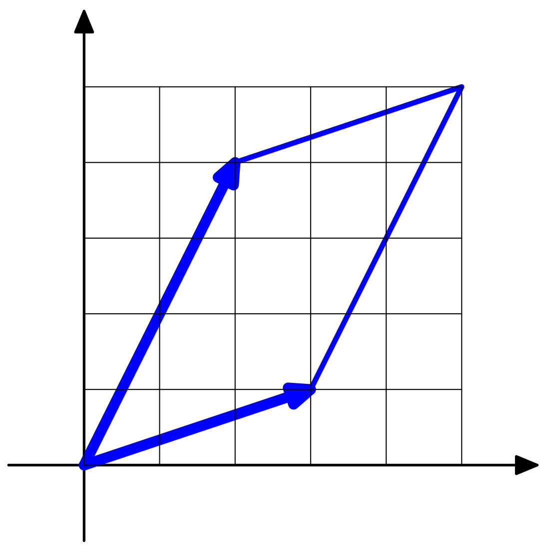
8.4 The Meaning of the Determinant: Area in the Plane

Let us consider an example in dimension 2 first. The following parallelogram is easily seen to have area 10 by chopping it into various triangles (two vertical cuts suffice yielding two triangles and an "easy" parallelogram).



Here, we shall demonstrate a different method that generalizes to higher dimensions. We will change the parallelogram and keep track of changes to its volumes along the way. In the end, we will obtain the unit square whose volume, we know, is 1. Row-operations will be our friend once again.

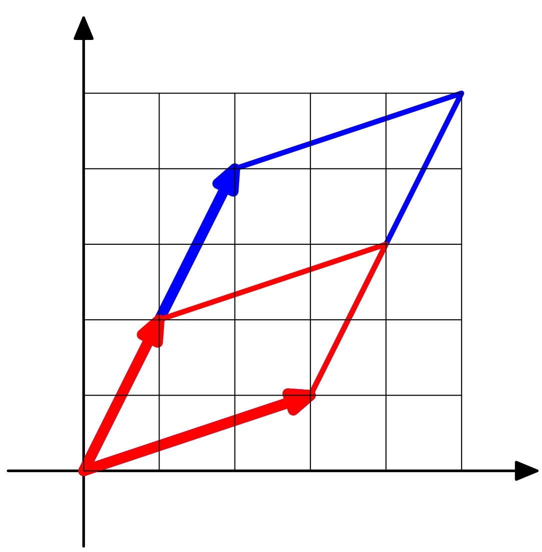
Note that the parallelogram above is spanned by two vectors: $\begin{bmatrix}3 & 1\end{bmatrix}$ and $\begin{bmatrix}2 & 4\end{bmatrix}$. We put those as row-vectors in one matrix:



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

Now, we row-reduce. Let us see what happens:

$\frac{1}{2}R_2 \to R_2$



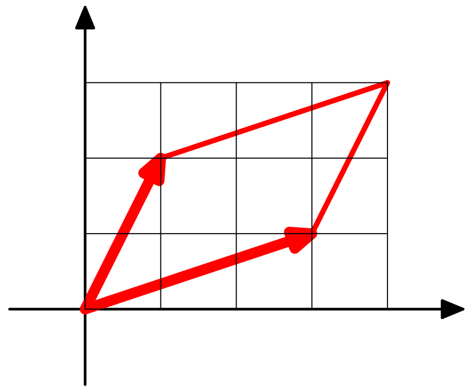
$$\begin{bmatrix}3 & 1\\ 2 & 4\end{bmatrix}$$

$$\tfrac{1}{2}R_2 \to R_2$$

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

Clearly, the area is cut in half.

$R_1 \leftrightarrow R_2$



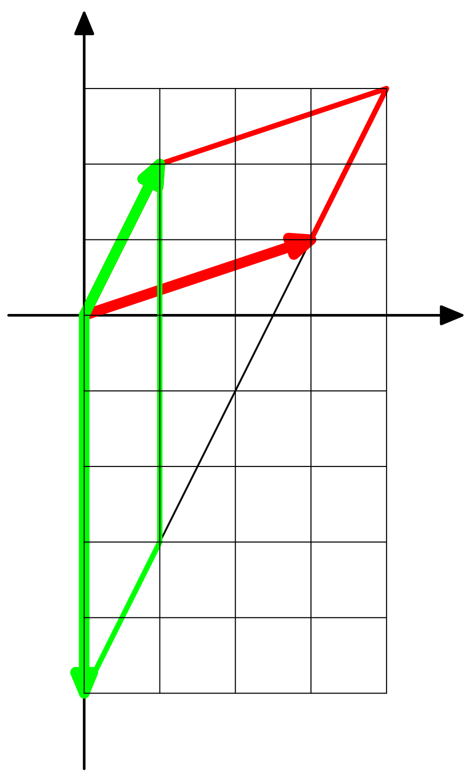
$$\begin{bmatrix} 3 & 1 \\ 1 & 2 \end{bmatrix}$$

$R_1 \leftrightarrow R_2$

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

The area does not change, in fact, the parallelogram stays the same, just the order of the defining vectors changed.

$3R_1 + R_2 \rightarrow R_2$



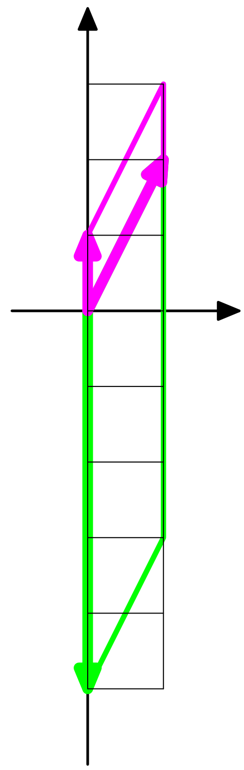
$$\begin{bmatrix} 1 & 2 \\ 3 & 1 \end{bmatrix}$$

$3R_1 + R_2 \rightarrow R_2$

$$\begin{bmatrix} 1 & 2 \\ 0 & -5 \end{bmatrix}$$

All that happens is that the parallelogram is sheared: the area does not change.

$-\frac{1}{5}R_2 \to R_2$



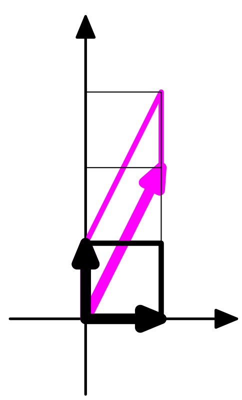
$$\begin{bmatrix} 1 & 2 \\ 0 & -5 \end{bmatrix}$$

$$-\tfrac{1}{5}R_2 \to R_2$$

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

Area changes by a factor of $\frac{1}{5} = |-\frac{1}{5}|$.

$-2R_2 + R_1 \to R_1$



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

$$-2R_2 + R_1 \to R_1$$

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

Again a row operation that does neither change area nor orientation.

Thus, overall, we changed orientation twice, cut the area in half once, and at one time, the area got multiplied by $\frac{1}{5}$. Thus, the original area is 10 times the final area, which is clearly 1. Hence, we started with an area of 10.

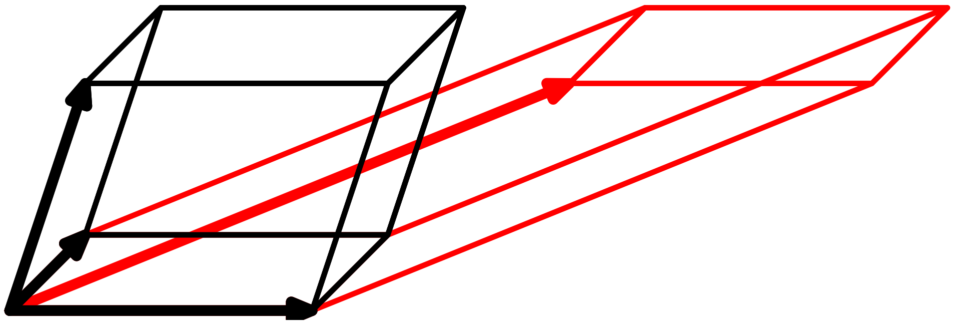
8.5 Three and More Dimensions

Looking back at the example above, we can formulate the underlying principles that allowed us to compute oriented area via row operations.

Observation 8.8. *Let A be an $n \times n$-matrix and let $P \subset \mathbb{R}^n$ be the parallelepiped spanned by the rows of A.*

1. *A row swap $R_i \leftrightarrow R_j$ does not change the volume of P.*
2. *A row operation $CR_j + R_i \to R_i$ does not change the volume of P.*

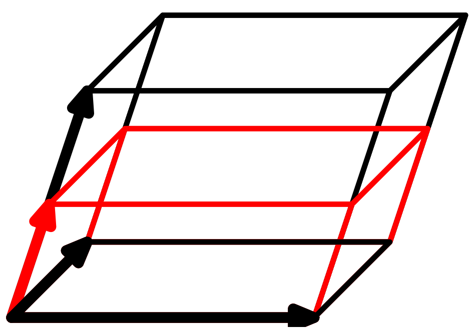
   This is apparent from the following picture:

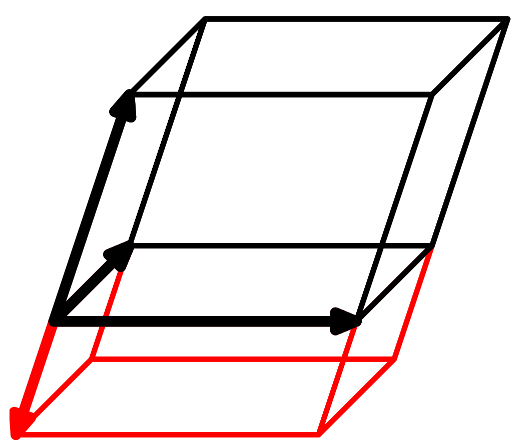
   Note how the base and height do not change. (One may think of this as shearing a deck of cards. The shape of the card is given by the parallelogram that is spanned by all rows but R_i. In the picture, it is the bottom.)

3. *A row operation $CR_i \to R_i$ multiplies the volume by $|C|$.*

   This is apparent from the following pictures. First the case $C > 0$.

   And now for $C < 0$.

q.e.d.

Remark 8.9. Compare this to Theorem 3.2. As you can see, the determinant obeys the very same rules as oriented volume. Since these rules enable us to compute oriented volumes (see the example in two dimensions), the determinant and oriented volume coincide (this also happened in the example above).

9 Inner Product Spaces

10 Eigenvalues and Eigenspaces

In this section, all vector spaces have finite dimension.

Definition 10.1. Let $\varphi : V \to V$ be an endomorphism (i.e., linear from a vector space to itself). For each scalar $\lambda \in \mathbb{F}$, we define the subspace

$$\mathrm{E}_{\varphi}(\lambda) := \ker(\varphi - \lambda\, \mathrm{id}_V) = \{\mathbf{v} \in V \mid \varphi(\mathbf{v}) = \lambda \mathbf{v}\}.$$

If $\mathrm{E}_{\varphi}(\lambda)$ contains a non-zero vector, we call λ an eigenvalue of φ with associated eigenspace $\mathrm{E}_{\varphi}(\lambda)$. The non-zero elements of $\mathrm{E}_{\varphi}(\lambda)$ are called eigenvectors.

Since every $r \times r$-matrix A describes a linear map

$$\begin{aligned} \mu_A : \mathbb{F}^r &\longrightarrow \mathbb{F}^r \\ \mathbf{x} &\mapsto A\mathbf{x} \end{aligned}$$

we apply the notions eigenvalue, eigenspace, and eigenvector to matrices. E.g., λ is an eigenvalue for A if there is a non-zero column vector $\mathbf{x} \in \mathbb{F}^r$ with $A\mathbf{x} = \lambda\mathbf{x}$.

Example 10.2. The kernel of φ is $\mathrm{E}_{\varphi}(0)$. Hence φ fails to be 1-1 if and only if 0 is an eigenvalue of φ.

10.1 Finding Eigenvalues: the Characteristic Equation

Definition 10.3. For a linear map $\varphi : V \to V$ where V has finite dimension, the determinant

$$\chi_{\varphi}(t) := \det(\varphi - t\, \mathrm{id}_V)$$

is a polynomial in t of degree $\dim(V)$. It is called the characteristic polynomial of φ.

The following observation yields a method for determining eigenvalues:

Observation 10.4. *For $\varphi : V \to V$ as above, a scalar $\lambda \in \mathbb{F}$ is an eigenvalue of φ if and only if it is a root of the characteristic polynomial, i.e, if and only if $\chi_{\varphi}(\lambda) = 0$. (This is immediate since a vanishing determinant detects a non-trivial kernel.)* **q.e.d.**

Definition 10.5. The equation $\chi(\varphi) = 0$ is called the characteristic equation of φ. Since eigenvalues of φ arise as roots of the characteristic polynomial, we can borrow the algebraic notion of multiplicities of roots: the multiplicity of an eigenvalue λ of φ is the the maximum exponent k so that $(t-\lambda)^k$ divides evenly into $\chi_{\varphi}(t)$.

Example 10.6. Find the eigenvalues of the matrix

$$[\ldots]$$

Also, determine their multiplicities and find bases for the associated eigenspaces.

10.2 Dimensions of Eigenspaces

Proposition 10.7. *Let $\varphi : V \to V$ be an endomorphism. Let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ be eigenvectors with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r$. Assume that for each eigenvalue, the corresponding eigenvectors in the above list are linearly independent. Then the list as a whole is a linearly independent set of vectors.*

Proof. We use induction on r.

Induction start: For $r = 1$, we have only a single vector $\mathbf{v}_1$. We need to show that this vector is non-zero (that is linear independence for a single vector). However, eigenvectors are by definition non-zero.

Induction step: Assume the lemma is true for a list of r eigenvectors. We want to show that it also holds for lists of size $r+1$. So let $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r, \mathbf{v}_{r+1}$ be such a list of eigenvectors with eigenvalues $\lambda_1, \lambda_2, \dots, \lambda_r, \lambda_{r+1}$. Since we assume that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$ are linearly independent, it suffices to show that $\mathbf{v}_{r+1}$ is not contained in the span of $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_r$. So let us assume to the contrary that

$$r_1\mathbf{v}_1 + \cdots + r_r\mathbf{v}_r = \mathbf{v}_{r+1}.$$

To continue the argument, we fix a basis of V that extends the linearly independent list $\mathbf{v}_1, \dots, \mathbf{v}_r$ and all coordinates below are relative to this basis. In particular, we have

$$\mathbf{v}_{r+1} = \begin{bmatrix} r_1 \\ \vdots \\ r_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

We also can compute the coordinates of

$$\begin{aligned} \varphi(\mathbf{v}_{r+1}) &= \varphi(r_1\mathbf{v}_1 + \cdots + r_r\mathbf{v}_r) \\ &= r_1\varphi(\mathbf{v}_1) + \cdots + r_r\varphi(\mathbf{v}_r) \\ &= r_1\lambda_1\mathbf{v}_1 + \cdots + r_r\lambda_r\mathbf{v}_r \end{aligned}$$

and since $\varphi(\mathbf{v}_{r+1}) = \lambda_{r+1}$, we have

$$\lambda_{r+1}\begin{bmatrix} r_1 \\ \vdots \\ r_r \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \lambda_{r+1}\mathbf{v}_{r+1} = \varphi(\mathbf{v}_{r+1}) = \begin{bmatrix} r_1\lambda_1 \\ \vdots \\ r_r\lambda_r \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

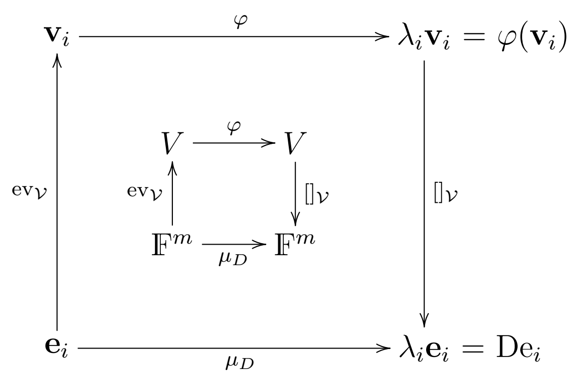
This shows that only those r_i can be non-zero, for which $\lambda_i = \lambda_{r+1}$. That, however, implies that $\mathbf{v}_{r+1}$ is in the span of those $\mathbf{v}_i$ with associates eigenvalue λ_{r+1}. The lemma assumes that in the given list, eigenvectors to the same eigenvalue are linear independent. So, we have a contradiction. **q.e.d.**

Corollary 10.8. $\sum_{\lambda} \dim(\mathrm{E}(\lambda)) \leq \dim(V)$.

Exercise 10.9. Show that $\dim(\mathrm{E}(\lambda))$ does not exceed the multiplicity of the eigenvalue λ. Hint: choose a basis for V that extends a basis for $\mathrm{E}(\lambda)$ and note that the matrix describing φ has a diagonal part. Note that this yields another proof of Corollary 10.8 since the degree of a polynomial is an upper bound for the sum of the multiplicities of its roots.

10.3 Diagonalizability and Diagonalization

If $\varphi : V \to V$ is a diagnonalizable endormorphism and $\mathcal{V} = \mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_m$ is a basis for V that consists of eigenvectors (with λ_i being the eigenvalue associated with $\mathbf{v}_i$), we have the following diagram:



Thus, the matrix D representing φ relative to a basis of eigenvectors is diagonal and features the eigenvectors along the diagonal.

Theorem 10.10. *Let $\varphi : V \to V$ be an endomorphism. Then, the following conditions are equivalent:*

1. *V has a basis relative to which φ is represented by a diagonal matrix.*
2. *V has a basis that consists entirely of eigenvectors for φ.*
3. *The inequality from Corollary 10.8 attains equality: $\sum_{\lambda} \dim(\mathrm{E}_{\varphi}(\lambda)) = \dim(V)$.*

Proof. If is clear that the first two conditions are equivalent and that the second implies the last. The remaining implication follows immediately from Proposition 10.7 since the basis for the different eigenspaces taken together yield a linearly independent set of size $\sum_{\lambda} \dim(\mathrm{E}_{\varphi}(\lambda)) = \dim(V)$ whence it is a basis. **q.e.d.**

Definition 10.11. An endomorphism satisfying one (and hence all) conditions from Theorem 10.10 is called diagonalizable.

Exercise 10.12. Show that an endomorphism is diagonalizable if its characteristic polynomial has $\dim(V)$ distinct roots. Note: this is not and "if and only if".

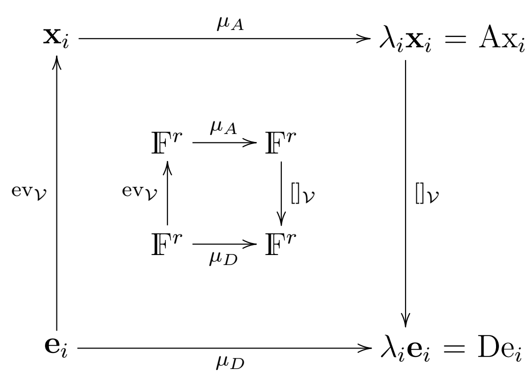
If the characteristic polynomial $\chi_\varphi(t)$ has $\dim(V)$ distinct roots, then φ is diagonalizable: each eigenspace has dimension at least 1 and there are $\dim(V)$ of them. Thus, their dimensions at up to (at least) $\dim(V)$. **q.e.d.**

10.4 Endomorphisms of Standard Vector Spaces

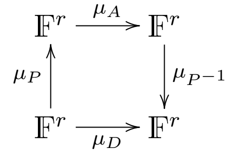
We know that every linear map from $\mathbb{F}^m$ to $\mathbb{F}^n$ is given as multiplication by a matrix of size $n \times m$ with coefficients in $\mathbb{F}$. For endomorphisms, the source and target vector space conincide. Hence, the matrix is a square matrix.

It is straight forward to specialize all previous notions and results to the case of an endomorphism that is given as left-multiplication by an $r \times r$ square matrix A: An eigenvector $\mathbf{x}$ with associated eigenvalue λ satisfies $A\mathbf{x} = \lambda\mathbf{x}$. The characteristic polynomial of A is $\chi_A(t) = \det(A - t\mathbb{I}_r)$ and its roots are the eigenvalues of A. The eigenspace $\mathrm{E}_A(\lambda)$ associated to λ is the null space of $A - \lambda\mathbb{I}_r$ and a basis for this eigenspace can be found by row-reduction.

Assuming that $\mathbb{F}^r$ has a basis $\mathcal{V}$ that consists of eigenvectors $\mathbf{x}_1, \ldots, \mathbf{x}_r$ for A, we get the following diagram:



Moreover, we know that the evaluation map $\mathrm{ev}_\mathcal{V}$ itself is given as multiplication by a matrix whose columns are exactly the basis vectors (this is because the evaluation map sends the standard basis vector $\mathbf{e}_i$ to $\mathbf{x}_i$ and that has to be equal to $P\mathbf{e}_i$). Thus, the diagram core becomes:



and we deduce: $P^{-1}AP = D$ or equivalently $A = PDP^{-1}$. Thus:

Proposition 10.13. *A diagonalizable matrix A is of the form $A = PDP^{-1}$ where D is a diagonal matrix. In this case, D features the eigenvalues for A along its diagonal and the columns of P give a basis for $\mathbb{F}^r$ that consists of corresponding eigenvectors.* **q.e.d.**

