A supplement to Treil

JIH

Version of: 13 April 2016

Throughout we use Treil to identify our text notes:

> Sergei Treil, *Linear Algebra Done Wrong* (9/7/2015 version),
> `https://www.math.brown.edu/~treil/papers/LADW/book.pdf`

As the title suggests, these notes are meant to supplement Treil rather than replace it. (In particular, these notes contain few examples.)

Throughout, when we refer to result (Theorem, Exercise, etc.) a.b.c of Treil, we mean result (Theorem, Exercise, etc.) b.c of Chapter a in Treil.

0 Chapter 0: Background

0.1 Sets and systems

We assume familiarity with the basics of set theory.

Given a set X, we will often discuss collections of its elements that are ordered and have repeated elements allowed. For instance

$$S, M, T, W, T, F, S$$

is such a collection of capital letters corresponding to the days of the week in the usual order.

Such a collection is formally an *ordered multiset*, but we will never use this term again. There are various names that can be used—for instance, *sequence*, *list*, *string*, and *word*. TREIL's preferred term is *system*; we shall keep to that, although we may also use *n-tuple*, when we want to emphasize that the system has exactly n entries; the example above is a 7-tuple of capital letters.

We usually write the system as above, separated by commas. We may also delimit it to avoid confusion, for instance when we need to give the system a name:

$$\mathcal{D} = [\![S, M, T, W, T, F, S]\!]$$

or

$$\mathcal{D} = (S, M, T, W, T, F, S)\,.$$

This last notation is particularly helpful when we we wish to think of the system as a *row vector* with entries from X.

Systems are typically indexed by some subset of the integers:

$$\mathcal{B} = [\![\mathbf{b}_1, \ldots, \mathbf{b}_j, \ldots, \mathbf{b}_n]\!] = [\![\mathbf{b}_j,\ 1 \le j \le n]\!]\,.$$

The number of elements n in a system (n-tuple) is its *length* or its *size*.

An *ordered set* is a special sort of system—one in which none of its elements is repeated in the system. Both $[\![T, i, g, e, r, s]\!]$ and $[\![Y, a, n, k, e, e]\!]$ are systems with six elements, but only the first is an ordered set. As e is repeated in the second system, it is not an ordered set.

To each system $\mathcal{X} = [\![x_1, \ldots, x_j, \ldots, x_n]\!]$ there is a unique associated set $\{\mathcal{X}\} = \{\, x_j \mid 1 \le j \le n \,\}$. Its cardinality is at most n, equality occurring precisely when $\mathcal{X}$ is an ordered set. Conversely, to each set $Y = \{\, y_i \mid 1 \le i \le m \,\}$ (with $y_i \ne y_j$ for $i \ne j$) there are $m!$ systems $\mathcal{Y}$ with $\{\mathcal{Y}\} = Y$.

Subsets of sets are familiar. There is a related concept for systems. If

$$\mathcal{B} = [\![\mathbf{b}_1, \ldots, \mathbf{b}_j, \ldots, \mathbf{b}_n]\!] = [\![\mathbf{b}_j,\ 1 \le j \le n]\!]\,.$$

is a system of length m, then a *subsystem* $\mathcal{C}$ is a system of length $m\,(\le n)$

$$\mathcal{C} = [\![\mathbf{b}_{c_1}, \ldots, \mathbf{b}_{c_i}, \ldots, \mathbf{b}_{c_m}]\!] = [\![\mathbf{b}_{c_i},\ 1 \le i \le m]\!]\,.$$

with $c_i < c_{i+1} \leq n$ for $1 \leq i < m$. We continue with the set theoretic notation, writing $\mathcal{C} \subseteq \mathcal{B}$. If $\mathcal{A} = [\![a_1, \ldots, a_m]\!]$ and $\mathcal{B} = [\![b_1, \ldots, b_n]\!]$ are systems of lengths m and n respectively, then we can concatenate them to get the new system

$$\mathcal{A} \cup \mathcal{B} = [\![a_1, \ldots, a_m, b_1, \ldots, b_n]\!]$$

of length $m + n$. We then have $\{\mathcal{A} \cup \mathcal{B}\} = \{\mathcal{C}\} = \{\mathcal{A}\} \cup \{\mathcal{B}\}$.

0.2 Mappings and functions

Given two sets A and B, a *function*, *map*, *transformation*, or *operator* f from A to B is something that associates to each member a of A (the *domain*) exactly one member b of B (the *target* or *codomain*). In notation,

$$\text{the function } f\colon A \longrightarrow B \text{ is given by } f(a) = b \text{ or } a \mapsto b\,.$$

The four names—function, map, transformation, and operator—for us mean the same thing. By convention different versions are used in different situations for various reasons.[1] For instance, we often use the term 'map' when no additional restrictions have been made.

The *range* (or *image*)

$$\operatorname{Ran} f = \{\, b \mid b = f(a), \text{ some } a \in A \,\}$$

is a subset of B but need not be all of B. While each $a \in A$ gives rise to a unique $b = f(a) \in B$, for a given $b \in B$ there may be no $a \in A$ with $f(a) = b$ or many. (We will return to these issues in the next subsection.)

Given two functions $f\colon A \longrightarrow B$ and $g\colon B \longrightarrow C$ their *composition* is the function $gf\colon A \longrightarrow C$ given by

$$gf(a) = g(f(a))$$

for all $a \in A$. The process can be iterated: if $h\colon C \longrightarrow D$ is a third function, then we have the composition $hgf\colon A \longrightarrow D$ given by

$$hgf(a) = h(g(f(a)))\,.$$

Composition of functions is associative:

$$(hg)f = hgf = h(gf)\,.$$

For each set X an important map (function) from X to itself is the *identity map*

$$\mathrm{Id}_X\colon X \longrightarrow X$$

given by

$$\mathrm{Id}_X(x) = x$$

for every $x \in X$. If $r\colon X \longrightarrow Y$ and $l\colon Y \longrightarrow X$ with $lr = \mathrm{Id}_X$, then r is a *right inverse* for l and l is a *left inverse* for r.

[1] So Treil uses the term 'linear operator' for vector space maps because his training is that of an analyst, while I am trained as an algebraist and am more likely to use 'linear transformation' in the same situation. They mean the same thing.

(0.1). LEMMA. *If $t\colon A \longrightarrow B$ has a left inverse $u\colon B \longrightarrow A$ and a right inverse $v\colon B \longrightarrow A$, then $u = v$.*

PROOF. $u = u(\mathrm{Id}_B) = u(tv) = (ut)v = (\mathrm{Id}_A)v = v$. □

The map t is an *inverse* (that is, *two-sided inverse*) provided it is both a left inverse and a right inverse.

(0.2). COROLLARY. *If the map $t\colon A \longrightarrow B$ has an inverse s, then s is unique.*

PROOF. If r and s are inverses for t, then r is a left inverse and s is a right inverse. Thus $r = s$ by the lemma. □

0.3 Isomorphism and solving equations

We introduce here the basic concept of an *isomorphism* of two mathematical objects. (See TREIL §1.6.3.) This is a formalism for saying that the two objects are essentially the same—they are "the same up to changing names" in the appropriate context.

For instance, two sets A and B with no further structure are "essentially the same set" precisely when they contain the same number of elements ($|A| = |B|$). Purely as sets

$$A = \{1, 2, 3, 4, 5\} \quad \text{and} \quad B = \{a, b, c, d, e\}$$

are the same although the names of the elements are different.

This set identification is formalized by finding a map t from A to B that accomplishes the "name change":

> *There is a map $t\colon A \longrightarrow B$ such that, for every $b \in B$, there is a unique $a \in A$ with $t(a) = b$.*

Such a map is called *bijective*. (The map is a *bijection*.) For instance, above we could take

$$t(1) = a\,,\ t(2) = b\,,\ t(3) = c\,,\ t(4) = d\,,\ t(5) = e\,.$$

At times it is easier to consider two related properties:

(i) *the map $t\colon A \longrightarrow B$ is* surjective *if, for every $b \in B$ there is at least one $a \in A$ with $t(a) = b$.*

(ii) *the map $t\colon A \longrightarrow B$ is* injective *if, for every $b \in B$ there is at most one $a \in A$ with $t(a) = b$.*

Clearly, a bijective map is precisely a map that is both surjective and injective.

In the following three problems, consider maps $f\colon A \longrightarrow B$, $g\colon B \longrightarrow C$, and their composition $gf\colon A \longrightarrow C$.

(0.3). PROBLEM.

(a) Prove that if f and g are surjective, then gf is surjective.

(b) Prove that if f and g are injective, then gf is injective.

(c) Prove that if f and g are bijective, then gf is bijective.

(0.4). PROBLEM.
(a) Prove that if gf is surjective, then g is surjective.
(b) Prove that if gf is injective, then f is injective.

(0.5). PROBLEM. *Give an example where g is surjective and f is injective, but gf is not surjective and not injective.*

We can think of this in terms of solving equations:

> *Given $b \in B$, how many solutions $x \in A$ are there to the equation $t(x) = b$?*

Solution of equations will be one of our main themes in linear algebra.

(0.6). PROPOSITION. *For the map $t\colon A \longrightarrow B$ the following are equivalent:*

(1) *t is surjective;*

(2) *for every $b \in B$ there is at least one solution $x \in A$ to $t(x) = b$;*

(3) *t has a right inverse.*

(0.7). PROPOSITION. *For the map $t\colon A \longrightarrow B$ the following are equivalent:*

(1) *t is injective;*

(2) *for every $b \in B$ there is at most one solution $x \in A$ to $t(x) = b$;*

(3) *t has a left inverse.*

(0.8). PROBLEM. *Prove Proposition (0.6).*

(0.9). PROBLEM. *Prove Proposition (0.7).*

In both of these, the equivalence of conditions (1) and (2) is essentially the definition. That (3) implies (1) is a consequence of a problem above. Therefore the real issue is to prove (1) implies (3). In your proof, notice that you will have choices to make, so that there will be more than one right (resp., left) inverse, unless t is a bijection. This observation is related to Lemma (0.1) above.

The two propositions immediately give:

(0.10). PROPOSITION. *For the map $t\colon A \longrightarrow B$ the following are equivalent:*

(1) *t is bijective;*

(2) *for every $b \in B$ there is a unique $x \in A$ to $t(x) = b$;*

(3) *t has an inverse.* □

(0.11). PROBLEM. *Let $t\colon A \longrightarrow B$ be a map with $|A| = |B|$, both finite. Prove that the following are equivalent:*

(1) *t is surjective;*

(2) *t is injective;*

(3) *t is bijective.*

0.4 Equivalence relations, representatives, and canonical forms

Consider a collection $\mathcal{S}$ of objects and a relation $\sim$ between members of $\mathcal{S}$. This relation is an *equivalence relation* on $\mathcal{S}$ provided it has three properties:

(i) (Reflexive) For all S in $\mathcal{S}$, $S \sim S$;

(ii) (Symmetric) For all S, T in $\mathcal{S}$, if $S \sim T$ then $T \sim S$;

(iii) (Transitive) For all S, T, U in $\mathcal{S}$, if $S \sim T$ and $T \sim U$, then $S \sim U$.

Of course, equality is the motivating example of an equivalence relation:

(i) (Reflexive) $S = S$;

(ii) (Symmetric) If $S = T$ then $T = S$;

(iii) (Transitive) If $S = T$ and $T = U$ then $S = U$.

But there are many more examples. In particular, isomorphism as described in the previous section is an equivalence relation on the class of all sets. (Exercise!)

Functions give us a great source for equivalence relations.

(0.12). LEMMA. *Let $f\colon A \longrightarrow B$ be a function. Define on A the relation*

$$a_1 \sim a_2 \iff f(a_1) = f(a_2)\,.$$

Then $\sim$ is an equivalence relation

The various equivalence classes are the *preimage* sets $f^{-1}(b)$ as b runs through B.

(0.13). PROBLEM. *Prove Lemma (0.12).*

Earlier we saw that bijections can be thought of as changes in names. Similarly equivalence relations can be thought of as assigning labels. The veterinary clinic might have as patients Boopie, Tiger, Snoopy, Ed, Tex, and Golden River, but for certain purposes it may be better to group them together as feline, feline, canine, equine, equine, equine. From a medical point of view, a horse is a horse.[2]

This reduction principle is important throughout mathematics. We discard distinctions not important to the situation at hand. For instance, let us say that two integers are congruent modulo 2 if their difference is a multiple of 2. This is an equivalence relation, and the two congruence classes are the even integers and the odd integers. Integer arithmetic induces a meaningful arithmetic on the two classes—the product of an odd integer and an even integer is always an even integer; the sum of an odd integer and an odd integer is always an even integer; and so forth.

[2] Of course, of course.

In practice, these large equivalence classes can be unwieldy. Instead we choose a *representative* for the class. For instance, among all finite sets with m elements, we might choose the representative set $\{1, 2, \ldots, m\}$. For the integers modulo 2 we could choose 2 to represent the even numbers and 17 to represent the odd numbers.

A particularly nice situation occurs when there is an algorithm that, for each element S of $\mathcal{S}$ produces a representative S' for the class of S that is canonical in the sense that S_1 and S_2 are equivalent if and only if $S_1' = S_2'$. In this case, S' is sometimes called a *canonical form*.

The sets $\{1, \ldots, m\}$ mentioned above are canonical for finite sets. For the integers modulo 2, we choose as canonical form for the integer z its remainder upon division by 2—that is, 0 for even integers and 1 for odd integers. The arithmetic described above then gives the set of representatives $\{0, 1\}$, written $\mathbb{Z}_2$, the following arithmetic structure, which we shall see below is that of a field with two elements:

$+$	0	1
0	0	1
1	1	0

$*$	0	1
0	0	0
1	0	1

.

0.5 Mathematical induction

Proof by induction has many versions, but its most basic setting is that of:

> **The Induction Principle.** *Consider a subset I of the positive integers $\mathbb{Z}^+$ that has the two properties:*
>
> (i) (Initialization step) $1 \in I$;
>
> (ii) (Induction step) *if $k \in I$, then $k + 1 \in I$.*
>
> *Then $I = \mathbb{Z}^+$, the set of all positive integers.*

Basic "proof by induction" involves some statement we wish to prove for all positive integers. We do this by verifying that the set I of all positive integers for which the statement is true satisfies the two steps of the Induction Principle.

Proof by induction is iterative in nature. It can be used to formalize proofs that include remarks such as "continuing in this manner" or "repeating this procedure."

Here is a classic example of proof by mathematical induction:

(0.14). THEOREM. *For every positive integer m we have*

$$1^2 + 2^2 + \cdots + (m-1)^2 + m^2 = \sum_{i=1}^{m} i^2 \;=\; \frac{m(m+1)(2m+1)}{6}\,.$$

PROOF. The proof is by induction on m. Let I be the set of positive integers m for which the identity is true. Then

(i) $1 \in I$ since $\sum_{i=1}^{1} i^2 = 1^2 = 1 = \frac{1(1+1)(2\cdot 1+1)}{6}$.

(ii) Assume $k \in I$; that is, $\sum_{i=1}^{k} i^2 = k(k+1)(2k+1)/6$. Consider $\sum_{i=1}^{k+1} i^2$. Then

$$\begin{aligned}\sum_{i=1}^{k+1} i^2 &= \left(\sum_{i=1}^{k} i^2\right) + (k+1)^2 \\ &= \frac{k(k+1)(2k+1)}{6} + (k+1)^2 \\ &= \frac{k(k+1)(2k+1) + 6(k+1)^2}{6} \\ &= \frac{(k+1)[k(2k+1) + 6(k+1)]}{6} \\ &= \frac{(k+1)(2k^2+7k+6)}{6} \\ &= \frac{(k+1)(k+2)(2k+3)}{6}\,.\end{aligned}$$

That is, $k+1 \in I$.

Therefore by the Induction Principle $I = \mathbb{Z}^+$, and the identity is true for all positive integers m. □

(0.15). PROBLEM. *Prove the canonical induction example:*
For every positive integer m we have

$$1 + 2 + \cdots + (m-1) + m = \sum_{i=1}^{m} i \;=\; \frac{m(m+1)}{2}\,.$$

(0.16). PROBLEM. *Consider the infinite sequence $(r_1, r_2, \dots)$ given by the recursion*

$$r_1 = a \quad \textit{and} \quad r_{i+1} = 3r_i + 2\,.$$

Prove that, for all k, we have $r_k = -1 + 3^{k-1}(a+1)$. REMARK. *This calculation was actually required in a recent research paper of mine.*

Usually the presentation is not quite as rigid as in the theorem. Here is a more typical version, which formalizes TREIL's "repeating this procedure."

(0.17). PROPOSITION. (TREIL's Proposition 1.2.8.) *Any finite generating system of a nonzero vector space contains a basis.*

PROOF. The proof is by induction on k, the number of vectors in a vector space generating system. As the spaces considered are nonzero, we have $k \geq 1$.

If a vector space has a generating system of size $k = 1$, then that generating system is a single nonzero vector $\mathbf{v}$, which is also linearly independent and so a basis. This initializes the induction.

Suppose now that the result is true for vector spaces with generating systems of size k. Consider a vector space with a generating system $\mathbf{v}_1, \dots, \mathbf{v}_{k+1}$ of size $k+1$. If the system is linearly independent, then it is a basis. If it is linearly dependent, then by Proposition 1.2.6 one of the vectors $\mathbf{v}_p$ can be written as a

linear combination of the others. When we delete $\mathbf{v}_p$ from the system, we are left with a generating system of size k. By induction, that system contains a basis. And clearly $\mathbf{v}_1, \ldots, \mathbf{v}_{k+1}$ contains that same basis.

Therefore the inductive step is valid, and so the proposition holds by induction. □

See Proposition (1.6) and Corollary (1.7) below for another proof of this result.

For a very typical proof by induction, see that for Lemma (1.20) given below.

There are many variants of the Induction Principle (including infinite versions). We give two helpful and elementary modifications.

> **The Induction Principle starting at c.** *Let c be an integer. Consider a subset I of $\mathbb{Z}_{\geq c}$, the set of all integers at least c, that has the two properties:*
>
> (i) (Initialization step) $c \in I$;
>
> (ii) (Induction step) *if $k \in I$, then $k+1 \in I$.*
>
> *Then I is equal to $\mathbb{Z}_{\geq c}$.*

The case $c = 1$ yields the original Induction Principle.

The version with $c = 0$ is often used. For instance, TREIL's Proposition 1.2.8 could have (and perhaps should have) been proven for all vector spaces (not just nonzero spaces) by induction on k starting at $k = 0$, since the only basis for a vector space $\{\mathbf{0}\}$ is the empty set $\emptyset$.

> **Complete Induction starting at c.** *Let c be an integer. Consider a subset I of $\mathbb{Z}_{\geq c}$ that has the property:*
>
> (a) (Complete induction step) *for $l \in \mathbb{Z}_{\geq c}$, if every integer i with $c \leq i < l$ is in I, then l is in I.*
>
> *Then I is equal to $\mathbb{Z}_{\geq c}$.*

(0.18). PROBLEM. *Prove that these last two forms of induction are equivalent to each other.* HINT: *If the subset I has (a), then $c \in I$ since trivially every integer i with $c \leq i < c$ is in I—there are no such i!*

(0.19). PROBLEM. *Consider the following induction proof that $k! \geq 2^k$ for all $k \geq 1$:*

Assume that the inequality is valid for k. Then

$$\begin{aligned} (k+1)! &= (k+1)k! \\ &\geq 2k! && \text{as } k \geq 1 \\ &\geq 2 \cdot 2^k && \text{by induction} \\ &\geq 2^{k+1} && \text{as desired.} \end{aligned}$$

(a) *What is wrong with this result and proof?*

(b) *Fix them.*

0.6 Fields

A field is a place where we can do arithmetic as usual. TREIL almost always uses the real numbers $\mathbb{R}$ (or later the complex numbers $\mathbb{C}$). You may also want to think of the rational numbers $\mathbb{Q}$ or the binary field $\mathbb{Z}_2$ (discussed in Section 0.4 above).

Formally a *field* is a set $\mathbb{F}$ together with two well-defined binary operations

$$+ : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F} \quad \text{and} \quad * : \mathbb{F} \times \mathbb{F} \longrightarrow \mathbb{F}$$

given by

$$(a, b) \mapsto a + b \quad \text{and} \quad (a, b) \mapsto a * b$$

and subject to the following axioms:

(1) (Additive commutativity) $a + b = b + a$ for all $a, b \in \mathbb{F}$;

(2) (Additive associativity) $a + (b + c) = (a + b) + c$ for all $a, b, c \in \mathbb{F}$;

(3) (Additive identity) there exists a element $0_{\mathbb{F}}$ in $\mathbb{F}$ such that $a + 0_{\mathbb{F}} = a$ for all $a \in \mathbb{F}$;

(4) (Additive inverses) for every $a \in \mathbb{F}$ there exists a $b \in \mathbb{F}$ such that $a + b = 0$;

(5) (Multiplicative identity) there exists a element $1_{\mathbb{F}} (\neq 0_{\mathbb{F}})$ in $\mathbb{F}$ such that $1_{\mathbb{F}} * a = a$ for all $a \in \mathbb{F}$;

(6) (Multiplicative associativity) $a * (b * c) = (a * b) * c$ for all $a, b, c \in \mathbb{F}$;

(7) (Left distributivity) $a * (b + c) = a * b + a * c$ for all $a, b, c \in \mathbb{F}$;

(8) (Right distributivity) $(a + b) * c = a * c + b * c$ for all $a, b, c \in \mathbb{F}$;

(9) (Mutliplicative inverses) for every $0_{\mathbb{F}} \neq a \in \mathbb{F}$ there exists a $b \in \mathbb{F}$ such that $a * b = 1_{\mathbb{F}}$;

(10) (Multiplicative commutativity) $a * b = b * a$ for all $a, b \in \mathbb{F}$.

The usual convention is to use juxtaposition ab in place $a * b$; for instance, (6) becomes $a(bc) = (ab)c$. Also we usually write 0 in place of $0_{\mathbb{F}}$ and 1 in place of $1_{\mathbb{F}}$ and hope that this does not cause confusion.

Subtraction and division are not defined. These are actually derived from addition and multiplication, using the inverses that are guaranteed. To "subtract" b from a, we add the additive inverse of b to a. To "divide" by nonzero b, we multiply by the multiplicative inverse of b.

The various axioms are presented in a small typeface to emphasize the fact that, while complicated, their message is simple: a field is a place where we can carry out the usual arithmetic operations of addition, subtraction, multiplication, and division satisfying the familiar rules (and subject to the familiar restrictions). Don't dwell on the axioms, just remember the message.

(0.20). PROBLEM. *From the axioms, prove:*

(a) *(Uniqueness of additive identity) $a + b = a$ for all $a \in \mathbb{F}$ if and only if $b = 0_{\mathbb{F}}$.*

(b) *(Uniqueness of additive inverse) For $a \in \mathbb{F}$, if $a + b = a + c$, then $b = c$. In particular, the additive inverse b of a guaranteed by (4) is uniquely determined and will be denoted $-a$.*

(c) $a * 0_{\mathbb{F}} = 0_{\mathbb{F}}$ *for all* $a \in \mathbb{F}$.

(d) $-a = (-1) * a$ *for all* $a \in \mathbb{F}$.

(e) *(Uniqueness of multiplicative identity)* $a * b = a$ *for all* $a \in \mathbb{F}$ *if and only if* $b = 1_{\mathbb{F}}$.

(f) *(Uniqueness of multiplicative inverse) For* $0_{\mathbb{F}} \neq a \in \mathbb{F}$, *if* $a * b = a * c$, *then* $b = c$. *In particular, the multiplicative inverse* b *of* a *guaranteed by (9) is uniquely determined and will be denoted* a^{-1}.

(g) *For* $a, b \in \mathbb{F}$, *if* $a * b = 0_{\mathbb{F}}$ *then* $a = 0_{\mathbb{F}}$ *or* $b = 0_{\mathbb{F}}$.

0.6.1 Matrix notation

For us $\mathrm{Mat}_{m,n}(X)$ will denote the set of all $m \times n$ matrices (rectangular arrays) with entries from the set X.

If A is the $m \times n$ matrix with entry $a_{i,j}$ in row i and column j then TREIL may write

$$A = (a_{i,j})_{i=1,\ j=1}^{m,\ n}$$

which we at times abbreviate to

$$A = (a_{i,j})_{i,j}\,.$$

We will be particularly interested in the *column* m*-tuples* from $X^m = \mathrm{Mat}_{m,1}(X)$ and the *row* n*-tuples* of $X_n = \mathrm{Mat}_{1,n}(X)$.

It will helpful at times to realize that the matrices of $\mathrm{Mat}_{m,n}(X)$ can be equally well thought of as column m-tuples with row vector entries from X_n and, especially, as row n-tuples with column vector entries from X^m:

$$A = \begin{pmatrix} a_{1,1} & a_{1,2} & a_{1,3} \\ a_{2,1} & a_{2,2} & a_{2,3} \end{pmatrix} = \begin{pmatrix} (a_{1,1}, a_{1,2}, a_{1,3}) \\ (a_{2,1}, a_{2,2}, a_{2,3}) \end{pmatrix} = \left(\begin{pmatrix} a_{1,1} \\ a_{2,1} \end{pmatrix} \begin{pmatrix} a_{1,2} \\ a_{2,2} \end{pmatrix} \begin{pmatrix} a_{1,3} \\ a_{2,3} \end{pmatrix} \right)\,.$$

If the $m \times n$ matrix A has $a_{i,j}$ as its row i and column j entry, then its *transpose* $A^\top$ is the $n \times m$ matrix that has $a_{i,j}$ as its row j and column i entry.

1 Chapter 1: Basic Notions

1.1 Vector spaces

We give a slightly modified version of TREIL's definition of a vector space:

Let $\mathbb{F}$ be a field. A *vector space V over $\mathbb{F}$* is a set V together with two well-defined maps, *vector space addition*

$$\oplus : V \times V \longrightarrow V$$

and *scalar multiplication*

$$\cdot : \mathbb{F} \times V \longrightarrow V$$

given by

$$(\mathbf{v}, \mathbf{w}) \mapsto \mathbf{v} \oplus \mathbf{w} \quad \text{and} \quad (\alpha, \mathbf{v}) \mapsto \alpha \cdot \mathbf{v}$$

and subject to the following axioms:

(1) (Additive commutativity) $\mathbf{v} \oplus \mathbf{w} = \mathbf{w} \oplus \mathbf{v}$ for all $\mathbf{v}, \mathbf{w} \in V$;

(2) (Additive associativity) $\mathbf{u} \oplus (\mathbf{v} \oplus \mathbf{w}) = (\mathbf{u} \oplus \mathbf{v}) \oplus \mathbf{w}$ for all $\mathbf{u}, \mathbf{v}, \mathbf{w} \in V$;

(3) (Additive identity) there exists a vector $\mathbf{0}_V$ such that $\mathbf{v} \oplus \mathbf{0}_V = \mathbf{v}$ for all $\mathbf{v} \in V$;

(4) (Additive inverses) for every vector $\mathbf{v} \in V$ there exists a $\mathbf{w} \in V$ such that $\mathbf{v} \oplus \mathbf{w} = \mathbf{0}_V$;

(5) (Multiplicative identity) $1_{\mathbb{F}} \cdot \mathbf{v} = \mathbf{v}$ for all $\mathbf{v} \in V$;

(6) (Multiplicative associativity) $\alpha \cdot (\beta \cdot \mathbf{v}) = (\alpha * \beta) \cdot \mathbf{v}$ for all $\alpha, \beta \in \mathbb{F}$ and $\mathbf{v} \in V$;

(7) (Left distributivity) $\alpha \cdot (\mathbf{v} \oplus \mathbf{w}) = \alpha \cdot \mathbf{v} \oplus \alpha \cdot \mathbf{w}$ in V for all $\alpha \in \mathbb{F}$ and $\mathbf{v}, \mathbf{w} \in V$;

(8) (Right distributivity) $(\alpha + \beta) \cdot \mathbf{v} = \alpha \cdot \mathbf{v} \oplus \beta \cdot \mathbf{v}$ in V for all $\alpha, \beta \in \mathbb{F}$ and $\mathbf{v} \in V$.

(1.1). PROBLEM. *From the axioms, prove:*

(a) *(Uniqueness of additive identity)* $\mathbf{v} \oplus \mathbf{w} = \mathbf{v}$ *for all* $\mathbf{v} \in V$ *if and only if* $\mathbf{w} = \mathbf{0}_V$.

(b) *(Uniqueness of additive inverse) For* $\mathbf{v} \in V$, *if* $\mathbf{v} \oplus \mathbf{x} = \mathbf{v} \oplus \mathbf{y}$ *then* $\mathbf{x} = \mathbf{y}$. *In particular, the additive inverse* $\mathbf{w}$ *of* $\mathbf{v}$ *guaranteed by (4) is uniquely determined and will be denoted* $-\mathbf{v}$.

(c) *For all* $a \in \mathbb{F}$, *we have* $a \cdot \mathbf{0}_V = \mathbf{0}_V$.

(d) *For all* $\mathbf{v} \in V$, *we have* $0_{\mathbb{F}} \cdot \mathbf{v} = \mathbf{0}_V$.

(e) *For all* $\mathbf{v} \in V$, *we have* $-\mathbf{v} = (-1) \cdot \mathbf{v}$.

We also use the terminology $\mathbb{F}$-*vector space* or even $\mathbb{F}$-*space* when referring to vector spaces over the field $\mathbb{F}$.

The notation emphasizes the fact that field addition and multiplication are not the same as scalar (vector space) multiplication and vector space addition. Nevertheless it is customary to denote both multiplications by juxtaposition; for instance, (6) becomes $\alpha(\beta\mathbf{v}) = (\alpha\beta)\mathbf{v}$. Similarly we use $+$ for both types of addition, so that (8) takes the form $(\alpha+\beta)\mathbf{v} = \alpha\mathbf{v} + \beta\mathbf{v}$. This is less cluttered but is open to misinterpretation, so care must be taken.

We typically write $\mathbf{0}$ in place of $\mathbf{0}_V$. As mentioned in the problem, the usual convention is to write $-\mathbf{v}$ for the additive inverse of $\mathbf{v}$ and always $-\mathbf{v} = (-1)\mathbf{v}$. We also write $\mathbf{w} - \mathbf{v}$ in place of $\mathbf{w} + (-\mathbf{v})$.

By Axiom (1) the conclusion of Axiom (3) could read $\mathbf{v} + \mathbf{0} = \mathbf{v} = \mathbf{0} + \mathbf{v}$, and similarly the conclusion of Axiom (4) could read $\mathbf{v} + \mathbf{w} = \mathbf{0} = \mathbf{w} + \mathbf{v}$.

It is not a coincidence that the axioms for a vector space are very similar to the axioms for a field (as given in Section 0.6). If $\mathbb{E}$ is a subfield of the field $\mathbb{F}$, then $\mathbb{F}$ has a natural structure as a vector space over $\mathbb{E}$. For instance, we often think of complex numbers as pairs of real numbers.

1.1.1 Examples

(i) We use $\mathrm{Mat}_{m,n}(\mathbb{F})$ to denote the vector space of all $m \times n$ matrices over the field $\mathbb{F}$. (See Section 0.6.1.) The corresponding notation in TREIL is $M_{m,n}$ with it understood that that entries come from the real field $\mathbb{R}$. The most important example of a vector space for us is $\mathbb{F}^m$ $(= \mathrm{Mat}_{m,1}(\mathbb{F}))$, the space of length n *column vectors*. The corresponding space $\mathbb{F}_n = \mathrm{Mat}_{1,n}(\mathbb{F})$ is that of *row vectors*.

(ii) A second type of important example of a vector space over $\mathbb{R}$ is the set $\mathbb{P}_n$ of all polynomials (with real coefficients) of degree at most n, provided with usual polynomial (function) addition and real multiplication by scalars (constant functions).

It is in fact possible, for any field $\mathbb{F}$, to define a vector space $\mathbb{P}_n(\mathbb{F})$ of polynomials of degree at most n with coefficients from $\mathbb{F}$. We will not give the definition here, although we observe that the vector spaces $\mathbb{P}_n(\mathbb{Q})$ and $\mathbb{P}_n(\mathbb{C})$ of rational and complex polynomials (respectively) with degree at most n have natural definitions similar to that of $\mathbb{P}_n = \mathbb{P}_n(\mathbb{R})$.

The vector space $\mathbb{P}_n$ can also be viewed as a special case of a *function space*. Let Φ be a collection of functions $f\colon X \longrightarrow W$, where X is some set and W is a vector space over the field $\mathbb{F}$. We can define function addition *pointwise*: while X does not itself have additive structure, the space W does. For every $x \in X$ and pair of functions $f, g \in \Phi$, the vectors $f(x)$ and $g(x)$ can be added together within the space W. Therefore we have the new function $f + g\colon X \longrightarrow \mathbb{F}$ given by

$$(f+g)(x) = f(x) + g(x), \text{ for } x \in X\,.$$

Similarly we have a pointwise definition of scalar multiplication: for $f \in \Phi$ and $a \in \mathbb{F}$ we get a new function af given by

$$(af)(x) = a(f(x)), \text{ for } x \in X\,.$$

For many choices of Φ, X, W, and $\mathbb{F}$, these operations give Φ the structure of a vector space over $\mathbb{F}$. $\mathbb{P}_n$ is the case where Φ consists of all real polynomial functions of degree at most n and $W = \mathbb{F} = \mathbb{R}$. We can also consider spaces such as $\mathrm{C}[0,1]$, the space of all continuous real functions defined on the interval $[0,1]$.

The matrix notation $A = (a_{i,j})_{i,j}$ shows that the matrix vector space $\mathrm{Mat}_{m,n}(\mathbb{F})$ can be thought of as a function space. It is the set of all functions $A\colon \{1,\dots,m\} \times \{1,\dots,n\} \longrightarrow \mathbb{F}$, where $A(i,j) = a_{i,j}$. In Sections 1.4 and 1.5.3 we shall see amother way in which the matrix space $\mathrm{Mat}(m,n)$ has a natural life as function space.

(1.2). Problem. *Prove that* $\mathrm{C}[0,1]$ *is a vector space over* $\mathbb{R}$.

(iii) As a rich source of examples of vector spaces, we introduce here the concept of *subspaces of vector spaces*. (See Treil §1.7.) These are the subsets W of the $\mathbb{F}$-vector space V that are themselves vector spaces over $\mathbb{F}$ for the vector addition and scalar multiplication that they inherit from V. If we examine the above axioms for a vector spaces we see that the subset W of V is a subspace provided it contains $\mathbf{0}$ and is closed under vector addition and scalar multiplication. (Care must be taken with the axiom guaranteeing additive inverses.)

Clearly the vector space V is a subspace of itself. Also the set $\{\mathbf{0}_V\}$ is a subspace of V, the *trivial space*. If $\mathbf{x}_1,\dots,\mathbf{x}_m$ is a system of vectors in the $\mathbb{F}$-vector space X, then its *span* in V is the subspace

$$\mathrm{Span}(\mathbf{x}_1,\dots,\mathbf{x}_m) = \{\sum_{j=1}^{m} \alpha_j \mathbf{x}_j \mid \alpha_j \in \mathbb{F}\}\,.$$

Treil uses the notations $\mathrm{span}\{\mathbf{x}_1,\dots,\mathbf{x}_m\}$ and $\mathcal{L}\{\mathbf{x}_1,\dots,\mathbf{x}_m\}$. I prefer the upper case Span, but Treil's curly brackets $\{\cdots\}$ have some virtue since the span only depends on the set $\{\mathbf{x}_1,\dots,\mathbf{x}_m\}$, not the particular ordered list $\mathbf{x}_1,\dots,\mathbf{x}_m$.[3]

1.1.2 Matrix notation

This is discussed in Section 0.6.1.

[3] We will not use Treil's notation $\mathcal{L}\{\mathbf{x}_1,\dots,\mathbf{x}_m\}$.

1.2 Linear combinations, bases

If $\alpha_1, \ldots, \alpha_j, \ldots \alpha_n$ is a system of elements in the field $\mathbb{F}$ and $\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n$ a system of vectors in the $\mathbb{F}$-vector space V, then the corresponding *linear combination* is the vector

$$\mathbf{v} = \alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \sum_{j_1}^{n} \alpha_j\mathbf{v}_j$$

of V, the α_i being the *coefficients* of the linear combination.

1.2.1 Generating and linearly independent systems

Consider the system $\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n$, and for arbitrary $\mathbf{v} \in V$ the equation

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{v}\,.$$

We have three important definitions:

(i) If for every $\mathbf{v} \in V$ there is at least one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to this equation, then the system of vectors is a *generating system* (or *spanning system*[4]).

(ii) If for every $\mathbf{v} \in V$, there is at most one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to this equation, then the system of vectors is a *linearly independent system.*

(iii) If for every $\mathbf{v} \in V$, there is exactly one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to this equation, then the system of vectors is a *basis* (or *base*).

By convention a trivial space $\{\mathbf{0}\}$ has a unique basis, namely, the empty set $\emptyset$. (This convention is particularly natural if we think of a basis as a minimal generating set, as in Corollary (1.7)(b) below.) The canonical basis $\mathbf{e}_1 \ldots, \mathbf{e}_m$ of $\mathbb{F}^m$ will be denoted $\mathcal{E}_m$, and the standard basis $1, t, \ldots, t^n$ of $\mathbb{P}_n$ will be denoted $\mathcal{S}_n$.

We have the important and immediate

(1.3). PROPOSITION. (TREIL's Proposition 1.2.7) *A system of vectors is a basis if and only if it is generating and linearly independent.* □

It may initially seem strange that a result, for which TREIL must work, is immediate for us. This is because our definitions are slightly different from those of TREIL. Our definitions match TREIL exactly for generating systems and bases, but our definition of linear independence is different, being (apparently) more restrictive than that used by TREIL:

[4] We will not use TREIL's term *complete*.

The system $\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n$ is *linearly independent* precisely when the only solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}$$

is the *trivial system* $\alpha_1 = \cdots = \alpha_j = \cdots = \alpha_n = 0$.

In any event, the system $\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n$ is *linearly dependent* if it is not linear independent. From our definition, this gives

for some $\mathbf{v} \in V$, there is more than one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{v}\,.$$

Equally well for TREIL, a system that is not linearly independent is *linearly dependent*. From his definition, this yields

there is a nontrivial solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}\,.$$

In this case, we may call the lefthand side of this equality a *nontrivial linear dependence*.

Correspondingly, the trivial system $\alpha_1 = \cdots = \alpha_j = \cdots = \alpha_n = 0$ is said to give a *trivial linear dependence* of the system.

The following lemma removes any confusion. It shows that the two concepts of linear dependence are equivalent, and hence (by the contrapositive) that the two definitions of linear independence are equivalent.

(1.4). LEMMA. *The following are equivalent:*

(1) *For some $\mathbf{v} \in V$, there is more than one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation*

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{v}\,.$$

(2) *There is more than one solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation*

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}\,.$$

(3) *There is a nontrivial solution system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_n$ to the equation*

$$\alpha_1\mathbf{v}_1 + \cdots + \alpha_j\mathbf{v}_j + \cdots + \alpha_n\mathbf{v}_n = \mathbf{0}\,.$$

PROOF. Parts (3) and (2) are equivalent, since the zero solution system $\alpha_i = 0$ always yields a trivial linear dependence. Furthermore (2) implies (1) by taking $\mathbf{v} = \mathbf{0}$.

It remains to show that (1) implies (2). Assume that $\beta_1, \ldots, \beta_j, \ldots, \beta_n$ and $\delta_1, \ldots, \delta_j, \ldots, \delta_n$ are two different solutions to the equation of (1) for $\mathbf{v}$. Then the homogeneous equation of (2) has the trivial solution $\alpha_i = 0$, for all i, but the homogeneous equation with righthand side $\mathbf{0} = \mathbf{v} - \mathbf{v}$ is also solved by the system $\gamma_1, \ldots, \gamma_j, \ldots, \gamma_n$ where we set $\gamma_i = \beta_i - \delta_i$, not all 0, as desired. □

(1.5). PROPOSITION. (TREIL's Proposition 1.2.6)

(a) *The system $\mathbf{v}_1, \ldots, \mathbf{v}_p$ is linearly dependent if and only if there is a k and a system $\beta_1, \ldots, \beta_p$ with $\mathbf{v}_k = \sum_{k \neq j=1}^{p} \beta_j \mathbf{v}_j$.*

(b) *More specifically, if $\mathbf{0} = \sum_{j=1}^{p} \alpha_j \mathbf{v}_j$ is a nontrivial linear dependence, then for every k with $\alpha_k \neq 0$ it is possible to find a system $\beta_1, \ldots, \beta_p$ with $\mathbf{v}_k = \sum_{k \neq j=1}^{p} \beta_j \mathbf{v}_j$.*

(c) *In the situation of* (b)*, the system $\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{v}_{k+1}, \ldots \mathbf{v}_p$ spans the same subspace of V as does $\mathbf{v}_1, \ldots, \mathbf{v}_p$.*

PROOF. (a) follows immediately from (b).

(b) ($\Longleftarrow$) With $\beta_k = -1$, $\mathbf{0} = \sum_{j=1}^{p} \beta_j \mathbf{v}_j$ is a nontrivial linear dependence.

($\Longrightarrow$) Let $\mathbf{0} = \sum_{j=1}^{p} \alpha_j \mathbf{v}_j$ be a nontrivial linear dependence by virtue of, say, $\alpha_k \neq 0$. Then $-\alpha_k \mathbf{v}_k = \sum_{k \neq j=1}^{p} \alpha_j \mathbf{v}_j$ and $\mathbf{v}_k = \sum_{k \neq j=1}^{p} \beta_j \mathbf{v}_j$ for $\beta_i = -\alpha_i (\alpha_k)^{-1}$.

(c) In this case, if $\mathbf{w} = \sum_{j=1}^{p} \gamma_j \mathbf{v}_j$ then

$$\begin{aligned} \mathbf{w} &= \sum_{j=1}^{p} \gamma_j \mathbf{v}_j = \gamma_k \mathbf{v}_k + \sum_{k \neq j=1}^{p} \alpha_j \mathbf{v}_j \\ &= \gamma_k \Big(\sum_{k \neq j=1}^{p} \beta_j \mathbf{v}_j \Big) + \sum_{k \neq j=1}^{p} \alpha_j \mathbf{v}_j \\ &= \sum_{k \neq j=1}^{p} (\gamma_k \beta_j + \alpha_j) \mathbf{v}_j \,. \end{aligned}$$ □

(1.6). PROPOSITION. *Let V be a finitely generated $\mathbb{F}$-space. For any linearly independent system $\mathcal{L}$ and any finite generating system $\mathcal{G}$, there is a subsystem $\mathcal{S}$ of $\mathcal{G}$ such that $\mathcal{L} \cup \mathcal{S}$ is a basis of V.*

PROOF. Let $\mathcal{S}$ be a subsystem of $\mathcal{G}$ of minimal size subject to V being generated by $\mathcal{L} \cup \mathcal{S}$. (Such an $\mathcal{S}$ exists since $\mathcal{G}$ is finite.) We claim that this generating system is linearly independent. Assume not. Then there is a nontrivial linear dependence among its members. In such a linear dependence, at least one nonzero coefficient must belong to a member of $\mathcal{S}$ since $\mathcal{L}$ itself is linearly independent. As noted in Proposition (1.5)(c), that element can be deleted from $\mathcal{S}$ with the remaining subsystem of $\mathcal{L} \cup \mathcal{S}$ still generating. This contradicts the minimality of $\mathcal{S}$. The contradiction proves that the generating system $\mathcal{L} \cup \mathcal{S}$ is also linearly independent and hence a basis, as claimed. □

(1.7). COROLLARY. *Let V be a finitely generated $\mathbb{F}$-space.*

(a) (TREIL's Proposition 1.2.8) *Every finite generating system contains a basis.*

(b) *Every minimal finite generating system is a basis.*

(c) (TREIL's Proposition 2.5.4) *Every linearly independent system in V is contained in a basis.*

(d) *Every maximal linear independent system is a basis.*

PROOF. The first two come from the case $\mathcal{L} = \emptyset$ of the proposition. The second two come from letting $\mathcal{L}$ be the system under discussion. □

Recall that part (a) of this corollary (that is, TREIL's Proposition 1.2.8) was also presented as Proposition (0.17) to provide an example of a proof by induction.

A special case of all of these is the following extremely important result:

(1.8). COROLLARY. *If V is a finitely generated $\mathbb{F}$-space, then it has a basis.* □

It is clear that every nonzero vector space has a generating system (for instance, the whole space) and a linearly independent system (for instance, any single vector), but it is not at all clear that every vector space has a basis. The above corollary guarantees that for finitely generated spaces.

In fact, every vector space has a basis. The corollaries (restated to remove finite generation) remain true for arbitrary vector spaces, but we will not pursue these extensions.

1.2.98 Sets and systems of vectors

If $\{\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n\}$ is a set of vectors in the $\mathbb{F}$-space V, then it is natural to consider the associated system $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$. Is it generating? linearly independent? a basis? Correspondingly, what do these properties for a system $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$ say about its underlying set $\{\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n\}$?

Recall that an ordered set is a system in which no element appears more than once.

Let $\mathcal{V}$ be an n-set in V (that is, a set containing n distinct vectors of V). We say that $\mathcal{V}$ is a *generating set* (or *spanning set*) if there is an ordered set (and so a system) $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$ that is a generating system with $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n\}$. Similarly the n-set $\mathcal{V}$ is *linearly independent* if there is an ordered set (system) $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$ that is a linearly independent system with $\mathcal{V} = \{\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n\}$.[5]

(1.9). LEMMA.

[5] We shall almost always be dealing with finite systems. For infinite systems and sets the same terminology is used, although then we must deal with infinite ordered sets.

(a) *If $\mathcal{V}$ is a generating set, then any ordering of its vectors gives a generating system.*

(b) *If $\mathcal{V}$ is a generating system, then its underlying set $\{\,\mathbf{v} \mid \mathbf{v} \in \mathcal{V}\,\}$ is a generating set.*

(c) *If $\mathcal{V}$ is a system and its underlying set $\{\,\mathbf{v} \mid \mathbf{v} \in \mathcal{V}\,\}$ is a generating set, then $\mathcal{V}$ is a generating system.*

(1.10). Lemma.

(a) *If $\mathcal{V}$ is a linearly independent set, then any ordering of its vectors gives a linearly independent system.*

(b) *If $\mathcal{V}$ is a linearly independent system, then it is an ordered set and its underlying set $\{\,\mathbf{v} \mid \mathbf{v} \in \mathcal{V}\,\}$ is a linearly independent set.*

(1.11). Problem.

(a) *Prove Lemma (1.9).*

(b) *Prove Lemma (1.10).*

Note the distinctions between the two results. The list

$$\mathcal{L} = [\![(1,0,0),(0,1,0),(0,0,1),(1,0,0)]\!]$$

is a generating system for $\mathrm{Mat}_{1,3}(\mathbb{F})$ with underlying set

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

which is a generating set. The set L is also linearly independent (any ordering of it is a basis), but the original system $\mathcal{L}$ is not linearly independent since it contains $(1,0,0)$ twice.

Since every linearly independent system is an ordered set of vectors, a basis is by definition a linearly ordered set of vectors that is both spanning (generating) and linearly independent. We sometimes abuse our terminology by referring to a set that is spanning and linearly independent as a basis. In this case, *any* ordering of the set gives a basis (different orderings giving different bases). So, for instance, a more common phrasing of Proposition (1.6) would be:

(1.12). Proposition. *Let V be a finitely generated $\mathbb{F}$-space. For any linearly independent set $\mathcal{L}$ and any finite generating set $\mathcal{G}$, there is a subset $\mathcal{S}$ of $\mathcal{G}$ such that $\mathcal{L} \cup \mathcal{S}$ is a basis of V.* □

1.2.99 Vector space isomorphism

Recall from Section 0.3 that an isomorphism exhibits two objects as basically same except the names of elements may have been changed. Here we are specifically interested in isomorphism of vector spaces (discussed in Treil's §1.6.3). A *vector space isomorphism* of the two $\mathbb{F}$-spaces V and W is a bijective map $T\colon V \longrightarrow W$

(1) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, for all $\mathbf{u}, \mathbf{v} \in V$;

(2) $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$, for all $\alpha \in \mathbb{F}$ and $\mathbf{v} \in V$.

So an isomorphism of the $\mathbb{F}$-spaces V and W is a set isomorphism of V and W that additionally respects the vector space operations that are defined on the two sets. The vector space W is essentially the same as the vector space V, only the names have been changed according to the map T. Vector space isomorphism formalizes natural feelings, such as the observation that there is no essential difference between row vectors and column vectors; see Lemma (1.15) below.

Two vector spaces V and W are *isomorphic* if there is an isomorphism $T\colon V \longrightarrow W$. When we think of an isomorphism as just a renaming of spaces that are essentially the same, it is clear that V is isomorphic to W if and only if W is isomorphic to V. Here is a more precise version of this observation:

(1.13). PROPOSITION. *Let $T\colon V \longrightarrow W$ be a vector space isomorphism. Especially T is a bijection of sets, so by Proposition (0.10) it has an inverse $S\colon W \longrightarrow V$ that is also a set isomorphism. In fact $S\colon W \longrightarrow V$ is an isomorphism of $\mathbb{F}$-vector spaces.*

PROOF. S is a bijection by definition.

Let $\mathbf{w}, \mathbf{x} \in W$ and $\alpha \in \mathbb{F}$. We must prove that

$$S(\mathbf{w}+\mathbf{x}) = S(\mathbf{w}) + S(\mathbf{x}) \quad \text{and} \quad S(\alpha\mathbf{w}) = \alpha S(\mathbf{w})\,.$$

Let $S(\mathbf{w}) = \mathbf{u}$ and $S(\mathbf{x}) = \mathbf{v}$, elements of V determined by the inverse set map S from W to V. Clearly $\mathbf{w} = TS(\mathbf{w}) = T(\mathbf{u})$ and $\mathbf{x} = TS(\mathbf{x}) = T(\mathbf{v})$. Now

$$\begin{aligned} S(\mathbf{w}+\mathbf{x}) &= S(T(\mathbf{u}) + T(\mathbf{v})) \\ &= S(T(\mathbf{u}+\mathbf{v})) \\ &= (ST)(\mathbf{u}+\mathbf{v}) \\ &= \mathbf{u}+\mathbf{v} \\ &= S(\mathbf{w}) + S(\mathbf{x})\,, \end{aligned}$$

where all the statements are at the set level except for the transition from the first line to the second where we have used the first property of T as vector space isomorphism. Next we use the second vector space isomorphism property of T to find

$$\begin{aligned} S(\alpha\mathbf{w}) &= S(\alpha T(\mathbf{u})) \\ &= S(T(\alpha\mathbf{u})) \\ &= (ST)(\alpha\mathbf{u})) \\ &= \alpha\mathbf{u} \\ &= \alpha S(\mathbf{w})\,, \end{aligned}$$

as desired. □

If there is some isomorphism from V to W, then we write $V \cong W$. An immediate consequence of the proposition is that isomorphism is symmetric: $V \cong W$ if and only if $W \cong V$. Clearly isomorphism is reflexive: the identity map is a vector space isomorphism of every vector space with itself. The next problem states that isomorphism is transitive. Therefore vector space isomorphism is an equivalence relation on the class of all $\mathbb{F}$-vector spaces.

(1.14). PROBLEM. *For $\mathbb{F}$-vector space isomorphisms $T_1 \colon V \longrightarrow W$ and $T_2 \colon U \longrightarrow V$, prove that the composition map $S = T_1 T_2 \colon U \longrightarrow W$ is a vector space isomorphism.*

(1.15). LEMMA. (TREIL's Proposition 1.2.7) *The vector space $\mathbb{F}^n$ (column vectors) and $\mathbb{F}_n$ (row vectors) are isomorphic $\mathbb{F}$-spaces, with the isomorphism given by the transpose map.*

PROOF. See Problem (1.19). □

TREIL's Remark 1.2.4 suggests the following extremely important result, repeated later in TREIL §1.6.3 as Example 2, and discussed in more detail in Section 2.8.1 below.

(1.16). THEOREM. *V has a basis $\mathbf{v}_1, \ldots, \mathbf{v}_m$ over $\mathbb{F}$ if and only if it is isomorphic as $\mathbb{F}$-vector space to $\mathbb{F}^m$.*

PROOF. ($\Longrightarrow$) For every $\mathbf{v}$ in V there is a uniquely determined system $\alpha_1, \ldots, \alpha_j, \ldots, \alpha_m$ from $\mathbb{F}$ with

$$\mathbf{v} = \alpha_1 \mathbf{v}_1 + \cdots + \alpha_j \mathbf{v}_j + \cdots + \alpha_m \mathbf{v}_m \,.$$

Define the map $T \colon V \longrightarrow \mathbb{F}^m$ by

$$T(\mathbf{v}) = (\alpha_1, \ldots, \alpha_j, \ldots, \alpha_m)^\top \,.$$

Then T is a vector space isomorphism.

($\Longleftarrow$) If $S \colon \mathbb{F}^m \longrightarrow V$ is a vector space isomorphism, then

$$\mathbf{v}_1 = S(\mathbf{e}_1), \ldots, \mathbf{v}_j = S(\mathbf{e}_j), \ldots, \mathbf{v}_m = S(\mathbf{e}_m)$$

is a basis of V. □

In the language of Section 0.4 every isomorphism class of finitely generated $\mathbb{F}$-vector spaces is represented by a column space $\mathbb{F}^m$ for some integer m.

(1.17). PROBLEM. *Prove that $\mathrm{Mat}_{m,n}(\mathbb{F})$ and $\mathrm{Mat}_{n,m}(\mathbb{F})$ are isomorphic $\mathbb{F}$-vector spaces via the transpose map.*

(1.18). PROBLEM.

(a) *Prove $\mathbb{P}_n$ is isomorphic to $\mathbb{R}^{n+1}$.*

(b) *Prove $\mathbb{P}_n$ is isomorphic to $\mathbb{R}_{n+1}$.*

(1.19). PROBLEM. *Prove that $\mathrm{Mat}_{m,n}(\mathbb{F})$, $(\mathbb{F}^m)_n$, and $(\mathbb{F}_n)^m$ are isomorphic $\mathbb{F}$-vector spaces.*

1.3 Linear Transformations

The concept of vector space isomorphism introduced in Section 1.2.99 was very natural. Here we discard the bijectivity requirement and reveal a powerful new topic.

A *linear transformation* $T\colon V \longrightarrow W$ of the $\mathbb{F}$-vector spaces V and W is a map from V to W with

(1) $T(\mathbf{u}+\mathbf{v}) = T(\mathbf{u}) + T(\mathbf{v})$, for all $\mathbf{u}, \mathbf{v} \in V$;

(2) $T(\alpha\mathbf{v}) = \alpha T(\mathbf{v})$, for all $\alpha \in \mathbb{F}$ and $\mathbf{v} \in V$.

Here V is the *domain* and W is the *target* or *codomain*.

Parts (1) and (2) of the definition above can be combined into the single "superposition" axiom:

$$T(\alpha\mathbf{u} + \beta\mathbf{v}) = \alpha T(\mathbf{u}) + \beta T(\mathbf{v}), \text{ for all } \alpha, \beta \in \mathbb{F} \text{ and all } \mathbf{u}, \mathbf{v} \in V.$$

This is the case $m = 2$ of the following important lemma:

(1.20). Lemma. $T(\sum_{j=1}^{n} \alpha_j \mathbf{v}_j) = \sum_{j=1}^{n} \alpha_j T(\mathbf{v}_j)\,.$

Proof. The proof is by induction on n. The case $n = 1$ is just (2) above.

Assume the result holds for $n-1$. Then by the superposition axiom and induction,

$$\begin{aligned} T\Big(\sum_{j=1}^{n} \alpha_j \mathbf{v}_j\Big) &= T\Big(\sum_{j=1}^{n-1} \alpha_j \mathbf{v}_j\Big) + \alpha_n \mathbf{v}_n \\ &= \Big(\sum_{j=1}^{n-1} \alpha_j T(\mathbf{v}_j)\Big) + \alpha_n \mathbf{v}_n \\ &= \sum_{j=1}^{n} \alpha_j T(\mathbf{v}_j)\,, \end{aligned}$$

as claimed. □

In particular (as observed in §1.3.3 on page 15 of Treil):

(1.21). Corollary. *A linear transformation $T\colon V \longrightarrow W$ is completely determined by its values on any generating system of V.* □

The case in which the system is a basis is particularly important.

(1.22). Theorem. *Let V and W be $\mathbb{F}$-spaces with $\mathbf{v}_1, \ldots, \mathbf{v}_n$ a basis of V and $\mathbf{w}_1, \ldots, \mathbf{w}_n$ any system of elements from W. Then the map*

$$T(\mathbf{v}_1) = \mathbf{w}_1, \ldots, T(\mathbf{v}_j) = \mathbf{w}_j, \ldots, T(\mathbf{v}_n) = \mathbf{w}_n$$

has a unique extension to a linear transformation $T\colon V \longrightarrow W$*, namely*

$$T\Big(\sum_{j=1}^{n} \alpha_j \mathbf{v}_j\Big) = \sum_{j=1}^{n} \alpha_j \mathbf{w}_j\,,$$

for all coefficient systems $\alpha_1, \ldots, \alpha_n$*.*

PROOF. The crucial observation is that the initial definition of T on the subset of the $\mathbf{v}_j$ has the given map as a well-defined extension to all of V. This works because the $\mathbf{v}_j$ form a basis. Every $\mathbf{v}$ has a unique expression as $\sum_{j=1}^{n} \alpha_j \mathbf{v}_j$; the system of coefficients $\alpha_1, \ldots, \alpha_n$ is uniquely determined by $\mathbf{v}$, so there is a well-defined image vector $\mathbf{w}$ for $\mathbf{v}$ given by $\sum_{j=1}^{n} \alpha_j \mathbf{w}_j$. And by Lemma (1.20) this extension is the only one that has a chance of being a linear transformation. It remains to check that this is indeed a linear transformation.

For $\mathbf{v}, \mathbf{w} \in V$ and $a, b \in \mathbb{F}$, if $\mathbf{v} = \sum_{j=1}^{n} \alpha_j \mathbf{v}_j$ and $\mathbf{w} = \sum_{j=1}^{n} \beta_j \mathbf{v}_j$ then $a\mathbf{v} + b\mathbf{w} = \sum_{j=1}^{n} (a\alpha_j + b\beta_j)\mathbf{v}_j$. Then

$$\begin{aligned} aT(\mathbf{v}) + bT(\mathbf{w}) &= a\sum_{j=1}^{n} \alpha_j \mathbf{v}_j + b\sum_{j=1}^{n} \beta_j \mathbf{v}_j \\ &= \sum_{j=1}^{n} a\alpha_j \mathbf{v}_j + \sum_{j=1}^{n} b\beta_j \mathbf{v}_j \\ &= \sum_{j=1}^{n} (a\alpha_j + b\beta_j)\mathbf{v}_j \\ &= T(a\mathbf{v} + b\mathbf{w})\,, \end{aligned}$$

as desired. □

We have discussed the basic definitions and properties of subspaces above. Here we have definitions of two important additional examples:

(a) If $A\colon X \longrightarrow Y$ is a linear transformation of $\mathbb{F}$-vector spaces then the *kernel* or *null space* of A is the subspace of X given by

$$\mathrm{Ker}(A) = \{\, \mathbf{x} \in X \mid A(\mathbf{x}) = \mathbf{0}_Y \,\}\,.$$

(b) If $A\colon X \longrightarrow Y$ is a linear transformation of $\mathbb{F}$-vector spaces then the *image* or *range* of A is the subspace of Y given by

$$\mathrm{Ran}(A) = \{\, \mathbf{y} \in Y \mid \text{there is } \mathbf{x} \in X \text{ with } A(\mathbf{x}) = \mathbf{y} \,\}\,.$$

(1.23). PROBLEM. *Prove that* $\mathrm{Ker}(A)$ *is a subspace of* X*.*

(1.24). PROBLEM. *Prove that* $\mathrm{Ran}(A)$ *is a subspace of* Y*.*

1.3.1 Examples of linear transformations

(i) Vector space isomorphisms are precisely the bijective linear transformations.

(ii) If $V = W$, then the *identity* linear transformation $I\colon V \longrightarrow V$ (at times written I_V) is given by $I(\mathbf{v}) = \mathbf{v}$ for all $\mathbf{v} \in V$.

(iii) The *trivial* (or *zero*) linear transformation $0_{V,W}\colon V \longrightarrow W$ is given by $0_{V,W}(\mathbf{v}) = \mathbf{0}_W$ for all $\mathbf{v} \in V$.

(iv) Differentiation $\mathbb{P}^n \longrightarrow \mathbb{P}^{n-1}$.

(v) Definite integration with domain $\mathrm{C}[0,1]$ and image $\mathbb{R}$: $f(x) \mapsto \int_0^1 f(t)dt$.

(vi) Rotation in $\mathbb{R}^2$.

(vii) Reflection in $\mathbb{R}^2$.

(viii) Projection of $\mathbb{F}^n$ onto $\mathbb{F}^m$ for $m < n$:
$(a_1, \ldots, a_m, a_{m+1}, \ldots, a_n)^\top \mapsto (a_1, \ldots, a_m)^\top$.

1.3.2 Matrix linear transformations and representation

Consider a linear transformation $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$. We shall call such a map a *matrix linear transformation*.

For each vector $\mathbf{e}_j$ in the standard basis $\mathcal{E}_n$ of $\mathbb{F}^n$ we set $\mathbf{a}_i = T(\mathbf{e}_j)$. The $m \times n$ matrix with $\mathbf{a}_j$ as its j^{th} column,

$$A = (\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n) \ ,$$

is then the *matrix representing* T.

If $\mathbf{x} = (x_1, \ldots, x_j, \ldots, x_n)^\top \in \mathbb{F}^n$ the linear transformation T has

$$T(\mathbf{x}) = x_1\mathbf{a}_1 + \cdots + x_j\mathbf{a}_j + \cdots + x_n\mathbf{a}_n \, .$$

(1.25). Lemma. $T(\mathbf{x}) = A\mathbf{x}$. □

If you are familiar with matrix multiplication, then the lemma is immediate from the previous displayed equation. Instead Treil takes the point of view that the lemma and displayed equation *define* matrix multiplication of the $m \times n$ matrix A by an $n \times 1$ matrix (vector) $\mathbf{x}$ with result the $m \times 1$ matrix (vector) $A\mathbf{x}\,(= T(\mathbf{x}))$.

At times we will write $[T] = A$ for the matrix representing the matrix linear transformation T. The equation of the lemma becomes

$$T(\mathbf{x}) = [T]\mathbf{x} \, ,$$

and we interpret the lemma to say that, for each matrix linear transformation T, there is a unique matrix $[T]$ that represents T (via matrix multiplication).

We will have much more to say about matrix representation in Section 2.8. Now we are content to observe that there is a one-to-one correspondence between matrix linear transformations and matrices. The direction $T \mapsto [T]$ of this correspondence has been described above. The other direction is given by the following lemma.

(1.26). LEMMA. *Let $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$. Then the map $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$ is a matrix linear transformation with $[T] = A$.*

(1.27). PROBLEM. *Prove this lemma.*

A common (but potentially confusing) convention is to identify the matrix A with its associated matrix linear transformation $\mathbf{x} \mapsto A\mathbf{x}$. That is, we sometimes write

$$A(\mathbf{x}) = A\mathbf{x}\,.$$

For the most part this will not cause problems, but care must be taken and the distinction should be remembered.

1.3.3 Linear transformations and generating sets

We discussed this in Corollary (1.21) above.

1.3.4 Conclusions

This material is discussed elsewhere.

1.4 Arithmetic properties of sets of linear transformations

Let $\mathcal{L}(V, W)$ be the set of all linear transformations from the $\mathbb{F}$-vector space V to the $\mathbb{F}$-vector space W. The set $\mathcal{L}(V, W)$ can be naturally interpreted as a function space from V to W, as in Section 1.1.1, providing it with the structure of an $\mathbb{F}$-vector space.

For scalar multiplication by elements of $\mathbb{F}$, for each $\alpha \in \mathbb{F}$ and each $T \in \mathcal{L}(V, W)$, we let the linear transformation $\alpha T \in \mathcal{L}(V, W)$ be given by

$$(\alpha T)(\mathbf{v}) = \alpha T(\mathbf{v}), \text{ for all } \mathbf{v} \in V\,.$$

This is a "pointwise" definition of the map αT, possible as the value $T(\mathbf{v})$ of the function T at the "point" $\mathbf{v}$ is a vector of W, a space that admits scalar multiplication by $\alpha \in \mathbb{F}$.

Similarly, for vector addition, when $S, T \in \mathcal{L}(V, W)$ the linear transformation $S + T \in \mathcal{L}(V, W)$ can be defined pointwise via

$$(S + T)(\mathbf{v}) = S(\mathbf{v}) + T(\mathbf{v}), \text{ for all } \mathbf{v} \in V\,.$$

(1.28). PROBLEM. *Let $V = \mathbb{F}^n$ and $W = \mathbb{F}^m$. Prove that $\mathcal{L}(V, W)$ is isomorphic to $\mathrm{Mat}_{m,n}(\mathbb{F})$.*

In certain circumstances it is also possible to "multiply" two linear transformations. (See TREIL §1.5.3.) If $T \in \mathcal{L}(V, W)$ and $U \in \mathcal{L}(W, X)$, where X is a third $\mathbb{F}$-space, then we can define $U \circ T = UT \in \mathcal{L}(V, X)$ pointwise by *composition* of functions:

$$(UT)(\mathbf{v}) = U(T(\mathbf{v})), \text{ for all } \mathbf{v} \in V\,.$$

It must be checked that each of αT, $S+T$, and UT defined above are genuinely $\mathbb{F}$-linear transformations (not just set maps).

(1.29). PROBLEM. *Check:*

(a) *αT is a linear transformation.*

(b) *$S + T$ is a linear transformation.*

(1.30). LEMMA. *UT is a linear transformation.*

PROOF. For $\mathbf{x}, \mathbf{y} \in V$ and $a, b \in \mathbb{F}$,

$$\begin{aligned} UT(a\mathbf{x} + b\mathbf{y}) &= U(T(a\mathbf{x} + b\mathbf{y})) \\ &= U(aT(\mathbf{x}) + bT(\mathbf{y})) \\ &= aU(T(\mathbf{x})) + bU(T(\mathbf{y})) \\ &= a(UT(\mathbf{x})) + b(UT(\mathbf{y}))\,. \square \end{aligned}$$

Having done that checking, we move on to verify that our algebraic operations on sets of linear transformations have various nice (and somewhat familiar properties). For instance, the first property follows from the associativity of function composition (discussed in Section 0.2). In each case we must assume that the appropriate domains and codomains are compatible as described above:

(1) Associativity: $A(BC) = (AB)C$;

(2) Distributivity: $A(B + C) = AB + AC$ and $(A + B)C = AC + AB$);

(3) Scalar commutativity: $A(\alpha B) = \alpha(AB) = (\alpha A)B$.

On the other hand, we cannot presume commutativity even with the appropriate compatibility:

(1.31). PROBLEM. *Find V and $A, B \in \mathcal{L}(V, V)$ with $AB \neq BA$.*

In the special case $V = W$, all compatibility conditions are valid. This gives the space $\mathcal{L}(V, V)$ a rich arithmetic structure.

1.5 Matrix multiplication

1.5.1 Matrix multiplication: definition

Recall from Section 1.3.2 the definition of the matrix product

$$A\mathbf{x} = x_1\mathbf{a}_1 + \cdots + x_j\mathbf{a}_j + \cdots + x_p\mathbf{a}_p$$

for $A \in \mathrm{Mat}_{m,p}(\mathbb{F})$ and column vector $(x_1, \ldots, x_j, \ldots, x_p)^\top = \mathbf{x} \in \mathbb{F}^p = \mathrm{Mat}_{p,1}(\mathbb{F})$. In the particular case $m = 1$, we have the product

$$(a_1, \ldots, a_j, \ldots, a_p)\mathbf{x} = \Big(\sum_{j=1}^{p} a_j x_j \Big),$$

a 1×1 matrix containing the *dot product* of the two p-tuples $\mathbf{a}$ and $\mathbf{x}$.

In the more general case $A\mathbf{x} = \mathbf{y}$ with $\mathbf{y}$ a column m-tuple, the i^{th} entry of $\mathbf{y}$ is the dot product of the i^{th} row of A with $\mathbf{x}$.

We now define the most general version of matrix multiplication. Let $A \in \mathrm{Mat}_{m,p}(\mathbb{F})$ and $B \in \mathrm{Mat}_{p,n}(\mathbb{F})$. Then

> *the product AB is the $m \times n$ matrix whose j^{th} column $A\mathbf{b}_j$ is the product of A and the j^{th} column $\mathbf{b}_j$ of B.*

Equivalently,

> *the product AB is the $m \times n$ matrix whose (i,j)-entry is the dot product of the i^{th} row of A and the j^{th} column of B.*

In both versions we find

$$(AB)_{i,j} = \sum_{k=1}^{p} a_{i,k} b_{k,j}$$

for $A = (a_{i,k})_{i,k}$ and $B = (b_{k,j})_{k,j}$.

It is important to understand that the product of two matrices with entries from $\mathbb{F}$ is defined if and only if the number of columns in the first matrix is equal to the number of rows in the second matrix.

1.5.2 Matrix multiplication: motivation

A nice consequence of Treil's definition of matrix multiplication in terms of linear transformations is that we are provided with motivation for the familiar but somewhat arbitrary looking formula above. The composition of matrix linear transformations is represented by the product of the individual representing matrices.

(1.32). Proposition. *If $T_2 \colon \mathbb{F}^n \longrightarrow \mathbb{F}^p$ and $T_1 \colon \mathbb{F}^p \longrightarrow \mathbb{F}^m$ are linear transformations, then*

$$[T_1 T_2] = [T_1][T_2]\,.$$

Proof. Let $A = [T_1]$ and $B = [T_2]$. Thus column j of B is $\mathbf{b}_j = T_2(\mathbf{e}_j)$, and the product AB is defined to have j^{th} column $A\mathbf{b}_j$. This is also the j^{th} column of $[T_1 T_2]$, since

$$(T_1 T_2)(\mathbf{e}_j) = T_1(T_2(\mathbf{e}_j)) = T_1(\mathbf{b}_j) = [T_1]\mathbf{b}_j$$

by Lemma (1.25). □

1.5.3 Arithmetic properties of sets of matrices

Much of this material was discussed in Section 1.4 above. In that section we noted that the set of functions $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ is an $\mathbb{F}$-vector space, while in Section 1.1.1 we saw this for $\mathrm{Mat}_{m,n}(\mathbb{F})$. In fact

(1.33). PROPOSITION. *The map $T \longrightarrow [T]$ gives an isomorphism of the $\mathbb{F}$-vector spaces $\mathcal{L}(\mathbb{F}^n, \mathbb{F}^m)$ and $\mathrm{Mat}_{m,n}(\mathbb{F})$.*

(1.34). PROBLEM. *Prove this lemma.*

In particular $\mathrm{Mat}_{m,n}(\mathbb{F})$ can be viewed as a function space, as promised in Section 1.1.1.

This proposition and Proposition (1.32) show that the algebraic properties of linear transformations discussed in Section 1.4 go over directly to matrices. That is, when appropriate:

(1) Associativity: $A(BC) = (AB)C$;

(2) Distributivity: $A(B + C) = AB + AC$ and $(A + B)C = AC + AB$);

(3) $A(\alpha B) = \alpha(AB) = (\alpha A)B$.

But often AB does not equal BA.

The above remarks illustrate the general fact that a result or concept for linear transformations can be immediately reinterpreted in the special case of matrix linear transformations to say something meaningful about matrices. The correspondence will be a direct application of Propositions (1.32) and (1.33).

This translation is usually so immediate that the appropriate matrix result will often not be specifically noted. This is the case in TREIL and also will be in these notes, except where the matrix results seem worthy of special note. See Lemma (1.47) and Theorem (1.52) below.

1.5.4 Transpose

Using the dot product version of matrix multiplication, we immediately find:

(1.35). LEMMA. $(AB)^\top = B^\top A^\top$. □

1.5.5 Trace

The *trace* of the $n \times n$ matrix $A = (a_{i,j})_{i,j}$ is the sum of its diagonal entries: $\mathrm{trace}(A) = \sum_{i=1}^{n} a_{i,i}$.

(1.36). LEMMA. *If B is $m \times n$ and C is $n \times m$, then* $\mathrm{trace}(BC) = \mathrm{trace}(CB)$.

(1.37). PROBLEM. *Prove Lemma (1.36).*

TREIL points out that this is a consequence of:

(1.38). PROBLEM.

(a) *Prove that for fixed A, the map $T\colon X \longrightarrow \mathrm{trace}(XA)$ is a linear transformation.*

(b) *Prove that for fixed A, the map $T\colon X \longrightarrow \mathrm{trace}(AX)$ is a linear transformation.*

1.5.99 Block matrix multiplication

This can be useful.

(1.39). PROPOSITION. *For $i, j, k \in \{1, 2\}$, let $A_{i,j}$ be an $m_i \times p_j$ matrix and $B_{j,k}$ be an $p_j \times n_k$ matrix (all over the same field $\mathbb{F}$). Then*

$$\begin{pmatrix} A_{1,1} & A_{1,2} \\ A_{2,1} & A_{2,2} \end{pmatrix} \begin{pmatrix} B_{1,1} & B_{1,2} \\ B_{2,1} & B_{2,2} \end{pmatrix} = \begin{pmatrix} A_{1,1}B_{1,1} + A_{1,2}B_{2,1} & A_{1,1}B_{1,2} + A_{1,2}B_{2,2} \\ A_{2,1}B_{1,1} + A_{2,2}B_{2,1} & A_{2,1}B_{1,2} + A_{2,2}B_{2,2} \end{pmatrix}.$$

PROOF. This is clear for the dot product case $m_1 = 1 = n_1$, $m_2 = 0 = n_2$. The general case then follows directly. □

More generally, any blocking of two matrices that allows all the necessary products of submatrices gives a valid block matrix multiplication. (This can be proved by inducting starting from the proposition.) For instance, TREIL's initial definition of matrix multiplication in Section 1.5.1 is the case in which $m_1 = m$ (and all other m_i are 0), $p_1 = n$, and $n_j = 1$ for $1 \leq j \leq n$.

1.6 Invertible linear transformations

1.6.1 Identity

We have introduced the identity linear transformation I_V in Section 1.3.1 above. The identity matrix linear transformation of $V = \mathbb{F}^n$ is represented by the $n \times n$ identity matrix $I_n = [I_{\mathbb{F}^n}]$, which has 1's on its diagonal and 0's off the diagonal. (We also may write the identity matrix as I with no subscript or as $I_{n,n}$.)

Identity matrices serve as a multiplicative identities for the matrix arithmetic of Section 1.5.3 in the strong sense that, for an $m \times n$ matrix A,

$$I_m A = A \quad \text{and} \quad A I_n = A\,.$$

1.6.2 Invertible transformation and isomorphisms

If $T\colon V \longrightarrow W$ is a linear transformation of $\mathbb{F}$-vector spaces, then the linear transformation $S\colon W \longrightarrow V$ is a *right inverse* of T if $ST = I_V$, a *left inverse* of T if $TS = I_W$, and a (2-sided) *inverse* of T if it is both a right and a left inverse.

Recall that, by Proposition (1.13), if the linear transformation T has an inverse as set map, then the inverse is itself a linear transformation.

(1.40). LEMMA. (TREIL's Theorem 1.6.1) *If T has both a right inverse R and a left inverse L, then $L = R$ is the unique inverse of T.*

PROOF. This is an immediate consequence of Lemma (0.1). □

Thus when T has an inverse, it is unique and is usually denoted T^{-1}.

The following two results should be compared with Propositions (0.6) and (0.7).

(1.41). THEOREM. *Let $A\colon X \longrightarrow Y$ be a linear transformation of $\mathbb{F}$-vector spaces. Then the following are equivalent:*

(1) *for every $\mathbf{b} \in Y$, there is at least one solution $\mathbf{x} \in X$ to the equation $A(\mathbf{x}) = \mathbf{b}$;*

(2) *A is an surjection;*

(3) $\operatorname{Ran}(A) = Y$*;*

(4) *for every generating system $\mathcal{G}$ in X, the image of $\mathcal{G}$ under A is a generating system in Y.*

PROOF. The first three parts are basically restatements of the definitions of surjection and range. As X always has generating systems, (4) implies (3). On the other hand, the image under A of any generating system $\mathcal{G}$ in X generates the range of A; so (3) implies (4). □

(1.42). THEOREM. *Let $A\colon X \longrightarrow Y$ be a linear transformation of $\mathbb{F}$-vector spaces. Then the following are equivalent:*

(1) *for every $\mathbf{b} \in Y$, there is at most one solution $\mathbf{x} \in X$ to the equation $A(\mathbf{x}) = \mathbf{b}$;*

(2) *A is an injection;*

(3) $\operatorname{Ker}(A) = \{\mathbf{0}_X\}$*;*

(4) *for every linearly independent system $\mathcal{I}$ in X, the image of $\mathcal{I}$ under A is a linearly independent system in Y.*

PROOF. The equivalence of the first two is the definition of injection. For the equivalence of parts (2) and (3),

$$A(\mathbf{x}_0) = A(\mathbf{x}_1) \iff A(\mathbf{x}_0 - \mathbf{x}_1) = \mathbf{0} \iff \mathbf{x}_0 - \mathbf{x}_1 \in \operatorname{Ker}(A).$$

Next (4) implies (3) since if $\mathbf{v}$ is a nonzero vector in $\operatorname{Ker}(A)$ then $[\![\mathbf{v}]\!]$ is a linearly independent system in X whose image in Y is linearly dependent. Finally, (3) implies (4) since if $\mathcal{I}$ is a linearly independent system in X whose image in Y is linearly dependent, then the coefficients of that linear dependence give a nontrivial linear combination of the elements of $\mathcal{I}$ that is in $\operatorname{Ker}(A)$. □

(1.43). PROBLEM. *Prove that if $[\![A(\mathbf{x}_1), \ldots, A(\mathbf{x}_n)]\!]$ is linearly independent, then $[\![\mathbf{x}_1, \ldots, \mathbf{x}_n]\!]$ is linearly independent.*

Our definition of vector space isomorphism and TREIL's are slightly different. In Section 1.2.99 we have defined a vector space isomorphism as a linear transformation that is bijective; in this section TREIL defines a vector space isomorphism as a linear transformation that is invertible as linear transformation. The next result, among other things, shows that these two definitions are equivalent.

(1.44). THEOREM. (Compare with TREIL's Theorems 1.6.6, 1.6.7, and 1.6.8.) *Let $A\colon X \longrightarrow Y$ be a linear transformation of $\mathbb{F}$-vector spaces. Then the following are equivalent:*

(1) *for every $\mathbf{b} \in Y$, there is a unique solution $\mathbf{x} \in X$ to the equation $A(\mathbf{x}) = \mathbf{b}$;*

(2) *A is a bijection;*

(3) $\mathrm{Ran}(A) = Y$ *and* $\mathrm{Ker}(A) = \{\mathbf{0}_X\}$;

(4) *for every basis $\mathcal{B}$ in X, the image of $\mathcal{B}$ under A is a basis in Y.*

(5) *the image of a basis $\mathcal{A}$ under A is a basis in Y;*

(6) *A has an inverse $B\colon Y \longrightarrow X$;*

(7) *A is an isomorphism.*

PROOF. The equivalence of the first four parts comes from the previous two theorems and our definition of vector space isomorphism from Section 1.2.99. Our definition of vector space isomorphism as a bijective linear transformation then says that these are equivalent to (7).

If A has an inverse, then it is bijective hence (6) implies (2). Conversely, if A is an isomorphism, then we saw in Proposition (1.13) that its set inverse is actually a linear transformation inverse; that is, (7) implies (6).

Clearly (4) implies (5), so we can finish the proof by showing that (5) implies (3). Assume (5). As the the image of $\mathcal{A}$ is a basis of Y, it spans Y hence $\mathrm{Ran}(A) = Y$. Suppose $\mathbf{x} = \sum_{i=1}^{n} \alpha_i \mathbf{a}_i$ is a nonzero vector in $\mathrm{Ker}(A)$ for distinct $\mathbf{a}_i$ in $\mathcal{A}$. Then the linearly independent system $[\![\mathbf{a}_1, \ldots, \mathbf{a}_n]\!]$ in X would have as its image in Y the linearly dependent system $[\![A(\mathbf{a}_1), \ldots, A(\mathbf{a}_n)]\!]$. This cannot happen by (5), so no such $\mathbf{x}$ exists and $\mathrm{Ker}(A) = \{\mathbf{0}_X\}$. □

(1.45). PROBLEM. *Let $A\colon X \longrightarrow Y$ with Y a finitely generated $\mathbb{F}$-space.*
(a) *Prove that A is an surjection if and only if A has an right inverse $B\colon Y \longrightarrow X$.*
(b) *Prove that A is an injection if and only if A has a left inverse $B\colon Y \longrightarrow X$.*

Again the problem should be compared with Propositions (0.6) and (0.7). The results of the problem remain true without the assumption of finite generation, but their proofs rely on the corresponding more general version of Corollary (1.7) which we proved only in the finitely generated case.

(1.46). LEMMA. (TREIL's Theorem 1.6.3) *If the linear transformations $S\colon V \longrightarrow W$ and $T\colon W \longrightarrow X$ both are invertible, then $TS\colon V \longrightarrow X$ is invertible and $(TS)^{-1} = S^{-1}T^{-1}$.*

PROOF.

$$(TS)(S^{-1}T^{-1}) = T(S(S^{-1}T^{-1})) = T((SS^{-1})T^{-1}) = T(I_W T^{-1}) = TT^{-1} = I_X\,.$$

$$(S^{-1}T^{-1})(TS) = S^{-1}(T^{-1}(TS)) = S^{-1}((T^{-1}T)S) = S^{-1}(I_W S) = S^{-1}S = I_V\,.$$

□

The relationship between invertible matrix linear transformations and invertible matrices is as expected. If $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$ then $B \in \mathrm{Mat}_{n,m}(\mathbb{F})$ is a *right inverse* of A if $AB = I_m$, a *left inverse* of A if $BA = I_n$, and a (2-sided) *inverse* of A if it is both a right and a left inverse, then it is unique (see below) and is denoted A^{-1}.

(1.47). Lemma.

(a) *Let $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$ be a matrix linear transformation. Then T is invertible if and only if $[T]$ is invertible. In this case $[T^{-1}] = [T]^{-1}$.*

(b) *Let $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$. Then A is an invertible matrix if and only if the linear transformation $T\colon \mathbb{F}^n \longrightarrow \mathbb{F}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$ is invertible. In this case T^{-1} is the linear transformation $S\colon \mathbb{F}^m \longrightarrow \mathbb{F}^n$ given by $S(\mathbf{y}) = A^{-1}\mathbf{y}$.*

(1.48). Lemma. *If the matrix A has both a right inverse R and a left inverse L, then $L = R$ is the unique inverse of A.*

(1.49). Problem.

(a) *Prove Lemma (1.47).*

(b) *Prove Lemma (1.48). (Compare with Lemma (0.1).)*

We do not (yet) have a linear transformation interpretation of the transpose, so the following lemma demands proof.

(1.50). Lemma. (Treil's Theorem 1.6.5) *If the matrix A has an inverse, then $A^\top$ has an inverse and $(A^\top)^{-1} = (A^{-1})^\top$.*

Proof. $I = I^\top = (AA^{-1})^\top = (A^{-1})^\top A^\top$ and $I = I^\top = (A^{-1}A)^\top = A^\top (A^{-1})^\top$. □

1.6.3 Isomorphism

This material has been discussed under Section 1.2.99 and the previous Section 1.6.2.

1.6.4 Invertibility and equations

This material has been discussed under Section 1.6.2.

Although the following results occur later in Treil, they naturally accompany the material of Section 1.6.2.

(1.51). Theorem. (Treil's Theorem 2.6.1) *Let $A\colon V \longrightarrow W$ be a linear transformation, and consider the equation*

$$A(\mathbf{x}) = \mathbf{b}\,.$$

If the equation has a specific solution $\mathbf{x}_0$, then the set of all solutions is the coset $\mathbf{x}_0 + \operatorname{Ker} A$.

In particular, $\operatorname{Ker} A$ *is the subspace of all solutions to the associated* homogeneous *equation*

$$A(\mathbf{x}) = \mathbf{0}\,.$$

PROOF. If $\mathbf{x}_0$ and $\mathbf{x}_1$ both solve the equation, then their difference $\mathbf{z} = \mathbf{x}_1 - \mathbf{x}_0$ solves the associated homogeneous equation and so is in $\operatorname{Ker}(A)$. On the other hand, if a vector $\mathbf{z}$ is in that kernel, then $\mathbf{x}_1 = \mathbf{x}_0 + \mathbf{z}$ solves the original equation. □

Of particular interest is the matrix version of this result.

(1.52). THEOREM. *Let* A *be an* $m \times n$ *matrix from* $\mathbb{F}$ *and consider the equation*

$$A\mathbf{x} = \mathbf{b}\,,$$

for a fixed $\mathbf{b} \in \mathbb{F}^m$*. If the equation has a specific solution* $\mathbf{x}_0 \in \mathbb{F}^n$*, then the set of all solutions* $\mathbf{x}$ *is the coset* $\mathbf{x}_0 + \operatorname{Ker} A$*.*

In particular, $\operatorname{Ker} A$ *is the subspace of all solutions in* $\mathbb{F}^n$ *to the associated homogeneous equation*

$$A\mathbf{x} = \mathbf{0}\,. \qquad \square$$

1.7 Subspaces

This material has been discussed under Section 1.1.1.

1.8 Application to computer graphics

This material is not part of the course.

2 Chapter 2: Systems of linear equations

The theme of this chapter is the solution of equations involving linear transformations and in particular systems of linear equations.

2.1 Different faces of linear systems

A system of m linear equations in n unknowns can be thought of and written as a single matrix equation

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

where the $m \times n$ matrix A, the *coefficient matrix*, contains the coefficients of the system, $\mathbf{x} = (x_1, \ldots, x_n)^\top$ is the vector of unknowns and $\mathbf{b} = (b_1, \ldots, b_m)^\top$ is the vector of constants (from the righthand side of the equations).

The basic observation is that for invertible E, the set of all $\mathbf{x}$ solving

$$EA\,\mathbf{x} = E\mathbf{b}$$

is the same as the solution set for the previous matrix equation. We spend a great deal of time in this chapter looking for matrices E that make this second equation easier to solve than the firts. Especially we seek E for which EA contains a lot of entries 0.

At times we consider the associated *augmented* matrix, which is the $m \times (n+1)$ matrix

$$(\,A\,|\,\mathbf{b}\,)\ ,$$

written in block matrix form. With the invertible matrix E, the augmented matrix is transformed into

$$(\,EA\,|\,E\mathbf{b}\,)\ ,$$

2.2 Solutions and echelon form

2.2.1 Elementary operations

There are three types of *elementary row operations* which can be carried out on the $m \times n$ matrix A:

(i) *Exchange*: exchange rows i and j;

(ii) *Scaling*: multiply row i by the nonzero constant r;

(iii) *Replacement*: add s times row j to row i.

The corresponding *elementary matrices* are:

(i) *Exchange*: $X_{i,j}$—exchange rows i and j of the identity matrix I_m;

(ii) *Scaling*: $S_i(r)$ for $r \neq 0$—multiply row i of I_m by the nonzero constant r;

(iii) *Replacement*: $R_{i,j}(s)$—add s times row j of I_m to row i.

Performing an elementary row operation on the matrix A is equivalent to left multiplying A by the appropriate elementary matrix E.

(2.1). PROPOSITION. *Let A be an $m \times n$ with entries from $\mathbb{F}$.*

(i) *$X_{i,j}A$ is the result of exchanging rows i and j of the matrix A;*

(ii) *$S_i(r)A$ is the result of multiplying row i of A by the nonzero constant r;*

(iii) *$R_{i,j}(s)A$ is the result of adding s times row j of A to row i.*

PROOF. This follows by direct calculation. □

Further calculation shows that inverses and transposes of elementary matrices are again elementary matrices.

(2.2). PROPOSITION.

(i) $X_{i,j} = X_{i,j}^{-1} = X_{j,i}^{\top}$.

(ii) $S_i(r) = S_i(r^{-1})^{-1} = S_i(r)^{\top}$ *for nonzero r.*

(iii) $R_{i,j}(s) = R_{i,j}(-s)^{-1} = R_{j,i}(s)^{\top}$. □

Although we will not need them for a while, the *elementary column operations* are found from the elementary row operations by replacing each instance of the word "row" with "column." Elementary column operations are carried out through *right* multiplication by elementary matrices:

(2.3). PROPOSITION. *Let B be an $n \times m$ with entries from $\mathbb{F}$.*

(i) *$BX_{i,j}$ is the result of exchanging columns i and j of the matrix B;*

(ii) *$BS_i(r)$ is the result of multiplying column i of A by the constant r;*

(iii) *$BR_{i,j}(s)$ is the result of adding s times column i of B to column j.*

PROOF. Set $B = A^{\top}$ in Proposition (2.1). □

2.2.2 Row reduction and Gaussian elimination

Let A be an $m \times n$ matrix with entries from $\mathbb{F}$. The *leading entry* in the nonzero row i of A is that nonzero entry $a_{i,j}$ with the smallest j—that is, furthest to the left. The matrix A is in *row echelon form* (usually abbreviated to *echelon form* and sometimes written *REF*) provided:

> *If $a_{i,j}$ is the leading entry in row i, then $a_{l,k} = 0$ for all $i \leq k \leq m$ and $1 \leq l \leq j$, except for $a_{i,j} \neq 0$.*

In particular, a zero matrix is in echelon form. The leading entries of a matrix in echelon form are the *pivot entries* or just *pivots* of the echelon form. The columns containing pivots are the *pivot columns* and the remaining columns are the *nonpivot columns* or *nonpivots*.

The process of *row reduction* or *Gaussian elimination* starts from an arbitrary matrix A and, by a sequence of elementary row operations (that is, by multiplying on the left by a sequence of elementary matrices) moves the matrix into row echelon form.

The algorithm is initialized by $A_1 = A$ and $k = 1$.

Step k: if there are no nonzero entries in row k or below in A_k, stop.

Otherwise, find the leftmost leading entry in one of these rows and, if necessary, exchange that row with row k of A_k. Then add multiples of the new row k to all rows from $k+1$ down to ensure that all those rows have 0 in that column. The resulting matrix is A_{k+1}.

Set k to $k+1$ and continue.

This clearly achieves the desired result. The algorithm is relatively practical, since its complexity is roughly cubic in the size of the matrix. (See page 92 of Treil for a more detailed discussion.) Although it is not necessary, it may be helpful to use various $S_k(r^{-1})$ to rescale the pivot entry r in nonzero row k to 1, even while still at Step k.

If E is the associated product of elementary matrices achieving row echelon form $R = EA$, then the two equations

$$A\mathbf{x} = \mathbf{b} \quad \text{and} \quad R\mathbf{x} = E\mathbf{b}$$

have the same solution sets of $\mathbf{x}$ since $R = EA$ for invertible E (by Proposition (2.2)). As R is echelon form, the second equation is easy to solve, using *back-solving* (that is, *back substitution*) if necessary. That is, solve from the bottom to the top of R.

Gaussian elimination will also be of theoretical use to us. The first example of this is:

(2.4). Theorem. *Every matrix can be written as the product of a sequence of elementary matrices and a matrix in echelon form.*

Especially, every square matrix can be written as the product of elementary matrices and an upper triangular matrix and as the product of a lower triangular matrix and elementary matrices.

Proof. If the echelon form of A is $R = EA$, then $E^{-1}R = A$. If A is square, then R is upper triangular. Also for square $B = A^\top$, we have $B = SF$ with $S = R^\top$ lower triangular and $F = (E^{-1})^\top = (E^\top)^{-1}$ a product of elementary matrices. □

2.2.3 Reduced row echelon form

A matrix may have many different row echelon forms. It would be nice to have one that is more canonical and in the bargain has many 0's.

The matrix A is in *reduced row echelon form* (usually abbreviated to *reduced echelon form* and sometimes written $RREF$) provided:

> *A is in echelon form and additionally each pivot value $a_{i,j}$ is* 1 *and, furthermore, $a_{l,j} = 0$ for all $1 \leq l < i$.*

This can be reached by elementary row operations (scaling and replacement) from any echelon form. This amounts to backsolving prior to attempting the solution of any equation.

If it is possible get from the matrix A to B in $\mathrm{Mat}_{m,n}(\mathbb{F})$ via a sequence of elementary row operations, then A and B are *row equivalent*. This gives an equivalence relation on the space of matrices, as defined in 0.4. Gaussian elimination allows us to find, for every matrix A, a matrix B that is row equivalent to A and in echelon form. In fact, reduced row echelon form is unique. That is, row reduced echelon form provides a canonical form under row equivalence.

2.3 Counting pivots

The matrix equation $A\mathbf{x} = \mathbf{b}$ is *consistent* if it has solutions, otherwise it is *inconsistent*.

The discussion of echelon form gives directly

(2.5). Lemma. *A system is inconsistent if and only if there is a pivot in the last column of an echelon form of its augmented matrix.* □

We mostly focus on echelon form for the coefficient matrix A. In relating the number of pivots to other matrix properties, the following trivial observation is crucial.

(2.6). Lemma. *A matrix in echelon form has at most one pivot in each row and at most one pivot in each column.* □

(2.7). Theorem. *Let A be an $n \times m$ matrix from $\mathbb{F}$, and let R be a row echelon form for A. We consider solutions $\mathbf{x} \in \mathbb{F}^m$ to the linear matrix equation $A\mathbf{x} = \mathbf{b}$ for the various $\mathbf{b} \in \mathbb{F}^n$.*

(a) *For all $\mathbf{b}$ the number of solutions $\mathbf{x}$ is greater than or equal to* 1 *if and only if R has a pivot in every row.*

(b) *For all $\mathbf{b}$ the number of solutions $\mathbf{x}$ is less than or equal to* 1 *if and only if R has a pivot in every column.*

(c) *For all $\mathbf{b}$ the number of solutions $\mathbf{x}$ is exactly equal to* 1 *if and only if R has a pivot in every row and every column.*

PROOF. (a) There are corresponding augmented matrices with pivots in the last column if and only if R has some zero rows at its bottom.

(b) The columns without pivots are free. If such columns exist then there are $\mathbf{b}$ admitting many solutions.

(c) This follows from the previous two parts. □

2.3.1 Dimension

(2.8). PROPOSITION. (TREIL's Proposition 2.3.1) *Let $\mathbf{v}_1, \ldots, \mathbf{v}_m$ be a system of vectors from $\mathbb{F}^n$, and construct the matrix A whose column j is $\mathbf{v}_j$. Let R be a row echelon form for A.*

(a) *The system $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is spanning if and only if R has a pivot in every row.*

(b) *The system $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is linearly independent if and only if R has a pivot in every column.*

(c) *The system $\mathbf{v}_1, \ldots, \mathbf{v}_m$ is a basis if and only if R has a pivot in every row and every column.*

PROOF. This comes directly from Theorem (2.7) and Section 1.2.1. □

(2.9). PROPOSITION. (TREIL's Proposition 2.3.5) *In $\mathbb{F}^n$ the size of a generating system is at least n.*

PROOF. This follows from Lemma (2.6) and Proposition (2.8)(a). □

(2.10). PROPOSITION. (TREIL's Proposition 2.3.2) *In $\mathbb{F}^n$ the size of a linearly independent system is at most n.*

PROOF. This follows from Lemma (2.6) and Proposition (2.8)(b). □

(2.11). PROPOSITION. (TREIL's Proposition 2.3.4) *In $\mathbb{F}^n$ the size of a basis is n.*

PROOF. This is a corollary to the last two propositions or to Proposition (2.8)(c). □

(2.12). PROPOSITION. (TREIL's Proposition 2.3.3) *If $\mathbf{v}_1, \ldots, \mathbf{v}_n$ is a basis of V, then all bases of V have size n.*

PROOF. As V is isomorphic to $\mathbb{F}^n$ by Theorem (1.16) and this isomorphism takes bases to bases by Theorem (1.44), the result is a corollary to the previous proposition. □

A vector space V with a finite basis has all bases of size n by Proposition (2.12). This number n is the *dimension* of V, denoted $\dim_{\mathbb{F}} V$. Our convention (and that of TREIL) is that any vector space not of finite dimension is said to have *infinite dimension*. (More precise statements can be made, but we do not pursue this.)

From Theorem (1.16) and Propositions (2.9) and (2.10) we get immediately:

(2.13). PROPOSITION. (TREIL's Proposition 2.5.3) *If* $\dim_{\mathbb{F}} V = n$*, then every generating system in* V *has greater than or equal to* n *elements.* □

(2.14). PROPOSITION. (TREIL's Proposition 2.5.2) *If* $\dim_{\mathbb{F}} V = n$*, then every linearly independent system in* V *has less than or equal to* n *elements.* □

2.3.2 Invertible matrices

(2.15). PROPOSITION. (See TREIL's Proposition 2.3.6) *For the matrix* $A \in \mathrm{Mat}_{m,n}(\mathbb{F})$ *with echelon form* R*, the following are equivalent:*

(1) A *is invertible;*

(2) *for all* $\mathbf{b} \in \mathbb{F}^m$ *the number of solutions* $\mathbf{x} \in \mathbb{F}^n$ *to* $A\mathbf{x} = \mathbf{b}$ *is exactly equal to* 1*;*

(3) R *has a pivot in each row and each column;*

(4) *the reduced row echelon form of* A *is* $I_m = I_n$*;*

(5) *the columns of* A *form a basis of* $\mathbb{F}^m$*.*

PROOF. The first two are equivalent by Theorem (1.44) (in the language of linear transformations). The second and third are equivalent by Theorem (2.7). The third is clearly equivalent to the fourth and is equivalent to the last by Proposition (2.8). □

In particular (4) gives an important result mentioned earlier:

(2.16). COROLLARY.

(a) (See TREIL's Corollary 2.3.7) *Invertible matrices must be square.*

(b) (See TREIL's Theorem 2.4.1) *Indeed every invertible matrix is a product of elementary matrices.*

PROOF. If A has I_n as row reduced echelon form, then there is a product E of elementary matrices with $EA = I_n$ whence $A = E^{-1}$ is a product of elementary matrices (by Lemma (1.46) and Proposition (2.2)). □

(2.17). PROPOSITION. (See TREIL's Proposition 2.3.8) *For* A *a square matrix, the following are equivalent:*

(1) A *is left invertible;*

(2) A *is right invertible;*

(3) A *is invertible.*

PROOF. Of course, if A is invertible then it is also left invertible and right invertible.

Now consider square matrices B and C with $BC = I$, so that C has left inverse B and B has right inverse C. Every equation $C\mathbf{x} = \mathbf{b}$ then has the unique solution $\mathbf{x} = B\mathbf{b}$, so by the previous proposition, C is invertible. But then, as in Lemmas (0.1) and (1.40), both B and C are invertible, being inverses. Let A first be B and then C to complete the proof. □

Lemma (1.46) says that, for A and B invertible, the product AB is also invertible. We have a partial converse.

(2.18). COROLLARY. *If A and B are square and their product AB is invertible, then A and B are invertible.*

PROOF. Let C be the inverse of AB. As $(AB)C = I = C(AB)$, A has the right inverse BC and B has the left inverse CA. □

2.4 Finding inverses

To calculate the inverse of invertible $n \times n$ matrix A, begin with the block $m \times 2m$ matrix

$$(\, A \,|\, I \,) \ .$$

Then for any matrix E we have

$$E\,(\, A \,|\, I \,) = (\, EA \,|\, E \,) \ .$$

In particular for invertible A the reduced row echelon form of $(\, A \,|\, I \,)$ is

$$E\,(\, A \,|\, I \,) = (\, EA \,|\, E \,) = (\, I \,|\, E \,) = \left(\, I \,|\, A^{-1} \,\right)$$

by Proposition (2.17), and the inverse matrix $A^{-1} = E$ has been found concretely as a product of elementary matrices.

2.5 Dimension

This material material has been discussed under Sections 1.2.1 and 2.3 above.

2.5.1 Completion to a basis

A nonconstructive version of TREIL's Proposition 2.5.4 was given in Corollary (1.7). We will return to a constructive version in Section 2.7.4.

2.6 General solution

This material has been discussed in part under Sections 1.6.4 and 2.2.2 above. What remains to be observed here is that in the process of completing the solution via backsolving, we have the following:

(2.19). PROPOSITION. *Consider a system of linear equations with matrix form*

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

that has solutions, as in Theorem (1.52).

The following numbers are equal:

(1) *The number of free variables in the general solution of the system.*

(2) *The number of nonpivot columns in an echelon form of A.*

(3) *The dimension of* $\ker A$. □

2.7 Fundamental subspaces and rank

Earlier we associated to any linear transformation $T\colon V \longrightarrow W$ two spaces:

$$\mathrm{Ker}(T) = \{\, \mathbf{v} \in V \mid T(\mathbf{v}) = \mathbf{0}_W \,\} \leq V \text{ and } \mathrm{Ran}(T) = \{\, T(\mathbf{v}) \mid \mathbf{v} \in V \,\} \leq W\,,$$

the kernel and range of T.

In Theorem (1.51) we decided that, for a linear transformation A and for each $\mathbf{b} \in \mathrm{Ran}(A)$, the set of solutions $\mathbf{x}$ to the equation $A(\mathbf{x}) = \mathbf{b}$ is the coset $\mathbf{x}_0 + \mathrm{Ker}\, A$, for an individual solution $\mathbf{x}_0$. This suggests a result of the shape

$$\dim \mathrm{Ran}(A) + \dim \mathrm{Ker}(A) = \dim V\,,$$

and we confirm this below in various forms.

For the matrix transformation of multiplication by an $m \times n$ matrix A, these are two of four *fundamental subspaces*:

- The *column space* of A, $\mathrm{CS}(A)$ is the subspace of $\mathbb{F}^m$ spanned by the columns of A. Its dimension is the *column rank* of A. In terms of the matrix linear transformation $\mathbf{x} \mapsto A\mathbf{x}$, we have $\mathrm{CS}(A) = \mathrm{Ran}(A)$.
- The (right) *null space* of A, $\mathrm{NS}(A)$ is the subspace of all $\mathbf{x} \in \mathbb{F}^n$ with $A\mathbf{x} = \mathbf{0}$. Its dimension is the (right) *nullity* of A. In terms of matrix linear transformations, $\mathrm{NS}(A) = \mathrm{Ker}(A)$.
- The *row space* of A, $\mathrm{RS}(A)$ is the subspace of $\mathbb{F}_n = \mathrm{Mat}_{1,n}(\mathbb{F})$ spanned by the rows of A. Its dimension is the *row rank* of A. We have $\mathrm{RS}(A) = (\mathrm{CS}(A^\top))^\top$.
- The *left null space* of A, $\mathrm{LNS}(A)$ is the subspace of of all $\mathbf{w} \in \mathbb{F}_m = \mathrm{Mat}_{1,m}(\mathbb{F})$ with $\mathbf{w}A = \mathbf{0}$. Its dimension is the *left nullity* of A. We have $\mathrm{LNS}(A) = (\mathrm{NS}(A^\top))^\top$.

Rather than these two final spaces, TREIL prefers to discuss their isomorphic transposed spaces $\mathrm{CS}(A^\top)$ and $\mathrm{NS}(A^\top)$.

A basic result is

(2.20). THEOREM. *Let E be an invertible $m \times m$ matrix, and A an $m \times n$ matrix.*

(a) $\mathrm{NS}(A) = \mathrm{NS}(EA)$.

(b) $\mathrm{RS}(A) = \mathrm{RS}(EA)$.

(c) $\dim_{\mathbb{F}} \mathrm{CS}(A) = \dim_{\mathbb{F}} \mathrm{CS}(EA)$.

(d) $\dim_{\mathbb{F}} \mathrm{LNS}(A) = \dim_{\mathbb{F}} \mathrm{LNS}(EA)$.

PROOF.

(a) $A\mathbf{x} = \mathbf{0}$ if and only if $EA\mathbf{x} = \mathbf{0}$, therefore $\mathrm{NS}(A) = \mathrm{NS}(EA)$.

(b) Set $\mathbf{y} = \mathbf{w}E$. $\mathbf{y}A = \mathbf{v}$ if and only if $\mathbf{w}EA = \mathbf{v}$, therefore $\mathrm{RS}(A) = \mathrm{RS}(EA)$.

(c) The map $\mathbf{w} \mapsto E\mathbf{w}$ is an isomorphism of $\mathrm{CS}(A)$ and $\mathrm{CS}(EA)$, therefore $\dim_{\mathbb{F}} \mathrm{CS}(A) = \dim_{\mathbb{F}} \mathrm{CS}(EA)$.

(d) The map $\mathbf{y} \mapsto \mathbf{y}E$ is an isomorphism of $\mathrm{LNS}(EA)$ and $\mathrm{LNS}(A)$, therefore $\dim_{\mathbb{F}} \mathrm{LNS}(A) = \dim_{\mathbb{F}} \mathrm{LNS}(EA)$. □

This immediately gives:

(2.21). COROLLARY. *Let R be an echelon form of A.*

(a) $\mathrm{RS}(A) = \mathrm{RS}(R)$.

(b) $\dim_{\mathbb{F}} \mathrm{CS}(A) = \dim_{\mathbb{F}} \mathrm{CS}(R)$. □

(2.22). THEOREM. (See TREIL's Theorem 2.7.1) *The column rank of A is equal to the row rank of A.*

PROOF. By Corollary (2.21) we only need prove this for matrices in echelon form. But in that case, both dimensions equal the number of pivots. □

We therefore may define $\mathrm{rank}(A)$, the *rank* of A, to be the common value of its column and row rank. TREIL defines $\mathrm{rank}(A)$ to the the column rank of A. Thus from his point of view this theorem states that

(2.23). COROLLARY. (TREIL 's Theorem 2.7.1) *A and $A^{\top}$ have the same rank.* □

In any event, the rank, column rank, and row rank of A are all the same.

2.7.1 Calculation

We wish to find explicitly a basis for each of the fundamental subspaces of the $m \times n$ matrix A with entries from $\mathbb{F}$. Let R be an echelon form of A with $R = EA$ for invertible E.

(I) $\mathrm{NS}(A)$: This is not new for us, since it is just the kernel of A; that is, the set of all solutions to $A\mathbf{x} = \mathbf{0}$. We solved such matrix equations easily by passing to the echelon form and noting that $R\mathbf{x} = \mathbf{0}$ has the same set of solutions $\mathbf{x}$.

(II) $\mathrm{RS}(A)$: By Corollary (2.21)(a) above, a basis for $\mathrm{RS}(A) = \mathrm{RS}(R)$ consists of the nonzero rows of the echelon form R for A.

(III) $\mathrm{CS}(A)$: The map $E\colon \mathbf{w} \mapsto E\mathbf{w}$ is an isomorphism of $\mathrm{CS}(A)$ and $\mathrm{CS}(R) = \mathrm{CS}(EA)$. As the pivot columns of R are a basis of $\mathrm{CS}(R)$, the images of these pivot columns under E^{-1} are columns of A that form a basis of $\mathrm{CS}(A)$. That is, the columns of A in the same positions as the pivot columns of R form a basis of $\mathrm{CS}(A)$.

(IV) $\mathrm{LNS}(A)$: This is a little trickier, and Treil does not really discuss it. However it turns out (exercise!) that if the zero rows of R are its last r rows, then the last r rows of E form a basis of $\mathrm{LNS}(A)$.

2.7.2 Explanation

Given above.

2.7.3 The rank theorem

(2.24). Theorem. (Rank plus nullity: Treil's Theorem 2.7.2) *Let A be an $m \times n$ matrix over $\mathbb{F}$.*

(a) $\dim_{\mathbb{F}} \mathrm{CS}(A) + \dim_{\mathbb{F}} \mathrm{NS}(A) = n$.

(b) $\dim_{\mathbb{F}} \mathrm{Ran}\, A + \dim_{\mathbb{F}} \mathrm{Ker}\, A = n$.

(c) $\dim_{\mathbb{F}} \mathrm{RS}(A) + \dim_{\mathbb{F}} \mathrm{LNS}(A) = m$.

(d) $\dim_{\mathbb{F}} \mathrm{Ran}\, A^{\top} + \dim_{\mathbb{F}} \mathrm{Ker}\, A^{\top} = m$.

Proof. For the first part, we note that the rank $\dim_{\mathbb{F}} \mathrm{CS}(A)$ is the number of pivot columns in the echelon form R while the nullity $\dim_{\mathbb{F}} \mathrm{NS}(A)$ is the number of nonpivot columns by Proposition (2.19).

The second part is the first, rephrased using linear transformation language. The last two parts are just the first two applied to the transpose $A^{\top}$. □

(2.25). Corollary. (Treil's Theorem 2.7.3) *Let A be an $m \times n$ matrix over $\mathbb{F}$. The equation*

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

has a solution $\mathbf{x}$ *for every* $\mathbf{b} \in \mathbb{F}^m$ *if and only if the dual equation*

$$A^\top \mathbf{y} = \mathbf{0}$$

has only the trivial solution $\mathbf{y} = \mathbf{0}$.

PROOF. Exercise. □

(2.26). PROBLEM. *Prove the corollary.*

2.7.4 Completion to a basis

We know by Corollary (1.7) that every linearly independent subset of a vector space $\mathbb{F}^n$ can be completed to a basis. Here is presented a concrete way of doing that. Namely, write the linearly independent m-set as the rows of an $m \times n$ matrix A. Then put A into an echelon form R. Then there are $n - m$ nonpivot ("free") columns. Add any set of $n - m$ vectors with the property that, for each nonpivot column, there is exactly one new vector whose leading entry is in that column. (For example, one can choose vectors that are all 0 except for a single 1 in a nonpivot column.) These vectors combined with the original m vectors gives $n = m + (n - m)$ vectors that (transposed) form a basis (since the row space of R is equal to the row space of E, and these vectors clearly complete the rows of R to a basis).

2.8 Matrix representation

2.8.1 Matrix representation of vector spaces

We recast our earlier observation Theorem (1.16). Let V be an $\mathbb{F}$-space with basis $\mathcal{A} = [\![\mathbf{a}_1, \ldots, \mathbf{a}_n]\!]$. Then for

$$\mathbf{v} = \sum_{j=1}^{n} \alpha_j \mathbf{a}_j$$

we write

$$[\mathbf{v}]_{\mathcal{A}} = (\alpha_1, \ldots, \alpha_n)^\top,$$

the *coordinate vector* of $\mathbf{v}$ for $\mathcal{A}$. Especially if $V = \mathbb{F}^n$ then $\mathbf{v} = [\mathbf{v}]_{\mathcal{E}}$ where $\mathcal{E}$ is the standard basis of $\mathbb{F}^n$.

(2.27). THEOREM. *The map*

$$\kappa_{\mathcal{A}} \colon \mathbf{v} \mapsto [\mathbf{v}]_{\mathcal{A}}$$

gives an isomorphism of V *and* $\mathbb{F}^n$. □

2.8.2 Matrix representation of linear transformations

Let $T\colon V \longrightarrow W$ be a linear transformation. Further let $\mathcal{A} = [\![\mathbf{a}_1, \ldots, \mathbf{a}_n]\!]$ be a basis of V and $\mathcal{B}$ a basis of W.

Define

$$[T]_{\mathcal{BA}}$$

to be the $m \times n$ matrix whose column j is

$$[T(\mathbf{a}_j)]_{\mathcal{B}}\,.$$

Then

(2.28). PROPOSITION. *$[T]_{\mathcal{BA}}$ is the unique matrix with $[T(\mathbf{v})]_{\mathcal{B}} = [T]_{\mathcal{BA}}[\mathbf{v}]_{\mathcal{A}}$ for all $\mathbf{v} \in V$.*

PROOF. Just as Theorem (2.27) is a more formal version of the earlier Theorem (1.16), so this result follows on from Lemma (1.25) and the remarks surrounding it.

Under the circumstances of the proposition, the map $S = \kappa_{\mathcal{B}} T \kappa_{\mathcal{A}}^{-1}$ is a linear transformation from $\mathbb{F}^n$ to $\mathbb{F}^m$ (for the appropriate dimensions of V and W). This is a matrix linear transformation, so by Lemma (1.25) there is a unique matrix $[S] = [\kappa_{\mathcal{B}} T \kappa_{\mathcal{A}}^{-1}]$ that represents it. As in Section 1.3.2, column j of $[S]$ is the vector

$$(\kappa_{\mathcal{B}} T \kappa_{\mathcal{A}}^{-1})(\mathbf{e}_j) = (\kappa_{\mathcal{B}} T)(\kappa_{\mathcal{A}}^{-1}(\mathbf{e}_j)) = (\kappa_{\mathcal{B}} T)(\mathbf{a}_j) = \kappa_{\mathcal{B}}(T(\mathbf{a}_j)) = [T(\mathbf{a}_j)]_{\mathcal{B}}\,.$$

That is, $[\kappa_{\mathcal{B}} T \kappa_{\mathcal{A}}^{-1}] = [T]_{\mathcal{BA}}$, as claimed. □

This result and its proof can be nicely expressed in terms of *commutative diagrams*. Specifically, given the linear transformation $T\colon V \longrightarrow W$ and bases $\mathcal{A}$ of V and $\mathcal{B}$ of W, the result says that the partial diagram of linear transformations

$$\begin{array}{ccc} W & \xleftarrow{\;T\;} & V \\ {\scriptstyle \kappa_{\mathcal{B}}}\big\downarrow & & {\scriptstyle \kappa_{\mathcal{A}}}\big\downarrow \\ \mathbb{F}^m & & \mathbb{F}^n \end{array}$$

completes uniquely to the commutative diagram of linear transformations

$$\begin{array}{ccc} W & \xleftarrow{\;T\;} & V \\ {\scriptstyle \kappa_{\mathcal{B}}}\big\downarrow & & {\scriptstyle \kappa_{\mathcal{A}}}\big\downarrow \\ \mathbb{F}^m & \xleftarrow[\;[T]_{\mathcal{BA}}\;]{} & \mathbb{F}^n \end{array}\,.$$

The proof makes use of Lemma (1.25), which is really just the special case of

matrix linear transformations:

$$\begin{array}{ccc} \mathbb{F}^m & \xleftarrow{S} & \mathbb{F}^n \\ \| & & \| \\ \mathbb{F}^m & \xleftarrow[[S]]{} & \mathbb{F}^n \end{array} .$$

We have the following important extension of the "rank plus nullity" Theorem (2.24).

(2.29). THEOREM. *Let $T\colon V \longrightarrow W$ be a linear transformation between finite dimensional vector spaces over $\mathbb{F}$. Then*

$$\dim_{\mathbb{F}} \operatorname{Ker} T + \dim_{\mathbb{F}} \operatorname{Ran} T = \dim_{\mathbb{F}} V .$$

PROOF. This is now immediate from Theorem (2.24)(a). □

In keeping with our earlier definitions, the *rank* of a linear transformation is the dimension of its range, while its *nullity* is the dimension of its kernel.

(2.30). PROPOSITION. *If additionally $S\colon W \longrightarrow X$ with $\mathcal{C}$ a basis of X, then*

$$[ST]_{\mathcal{CA}} = [S]_{\mathcal{CB}}[T]_{\mathcal{BA}} .$$

PROOF. By Proposition (2.28) the partial diagram

$$\begin{array}{ccc} X & \xleftarrow{ST} & V \\ {\scriptstyle \kappa_{\mathcal{C}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{A}}}\downarrow \\ \mathbb{F}^l & & \mathbb{F}^n \end{array}$$

completes uniquely to the commutative diagram

$$\begin{array}{ccc} X & \xleftarrow{ST} & V \\ {\scriptstyle \kappa_{\mathcal{C}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{A}}}\downarrow \\ \mathbb{F}^l & \xleftarrow[[ST]_{\mathcal{CA}}]{} & \mathbb{F}^n \end{array} .$$

On the other hand we can glue two commutative diagrams together to get a larger commutative diagram:

$$\begin{array}{ccccc} X & \xleftarrow{S} & W & \xleftarrow{T} & V \\ {\scriptstyle \kappa_{\mathcal{C}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{B}}}\downarrow & & {\scriptstyle \kappa_{\mathcal{A}}}\downarrow \\ \mathbb{F}^l & \xleftarrow[[S]_{\mathcal{CB}}]{} & \mathbb{F}^m & \xleftarrow[[T]_{\mathcal{BA}}]{} & \mathbb{F}^n \end{array} .$$

By the uniqueness of the first completion, we conclude

$$[ST]_{\mathcal{CA}} = [S]_{\mathcal{CB}}[T]_{\mathcal{BA}} . \quad \square$$

2.8.3 Change of coordinates for vector spaces

Assume $V = W$ and specialize Proposition (2.28) to the case $T = I_V$. Then we find

$$[\mathbf{v}]_{\mathcal{B}} = [I(\mathbf{v})]_{\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}}[\mathbf{v}]_{\mathcal{A}}\,.$$

That is, the matrix

$$[I]_{\mathcal{B}\mathcal{A}}$$

is the *change of coordinates* (or *change of basis* or *base change*) matrix for V, from the basis $\mathcal{A}$ to the basis $\mathcal{B}$.

As $[I]_{\mathcal{A}\mathcal{A}} = [I]_{\mathcal{B}\mathcal{B}} = I$, we have

$$[I]_{\mathcal{A}\mathcal{B}} = [I]_{\mathcal{B}\mathcal{A}}^{-1}\,.$$

For instance, if $V = \mathbb{F}^n$ then always $\mathbf{v} = [\mathbf{v}]_{\mathcal{E}}$ for the standard basis $\mathcal{E}$ of $\mathbb{F}^n$, so $[I]_{\mathcal{E}\mathcal{A}}$ is easy to find: its column j is $\mathbf{a}_j$. Then $[I]_{\mathcal{A}\mathcal{B}}$ can be calculated via

$$[I]_{\mathcal{A}\mathcal{B}} = [I]_{\mathcal{A}\mathcal{E}}[I]_{\mathcal{E}\mathcal{B}} = [I]_{\mathcal{E}\mathcal{A}}^{-1}[I]_{\mathcal{E}\mathcal{B}}\,.$$

2.8.4 Change of coordinates for linear transformations

Let $T\colon V \longrightarrow W$ with $\mathcal{A}$ and $\mathcal{C}$ bases of V and $\mathcal{B}$ and $\mathcal{D}$ bases of W. Then using change of coordinate matrices for V and W, we can "change coordinates" for T:

$$[T]_{\mathcal{D}\mathcal{C}} = [I]_{\mathcal{D}\mathcal{B}}[T]_{\mathcal{B}\mathcal{A}}[I]_{\mathcal{A}\mathcal{C}}\,.$$

2.8.5 Similarity of matrices

If above we take $V = W$, $\mathcal{A} = \mathcal{B}$, and $\mathcal{C} = \mathcal{D}$, then we find

$$[T]_{\mathcal{C}\mathcal{C}} = [I]_{\mathcal{C}\mathcal{B}}[T]_{\mathcal{B}\mathcal{B}}[I]_{\mathcal{B}\mathcal{C}} = [I]_{\mathcal{B}\mathcal{C}}^{-1}[T]_{\mathcal{B}\mathcal{B}}[I]_{\mathcal{B}\mathcal{C}} = Q^{-1}[T]_{\mathcal{B}\mathcal{B}}Q\,,$$

where $Q = [I]_{\mathcal{B}\mathcal{C}}$.

Two square matrices A and B are *similar* if there is an invertible matrix Q with

$$B = Q^{-1}AQ \quad \text{whence} \quad A = (Q^{-1})^{-1}BQ^{-1}\,.$$

In this case, Q may be thought of as a base change matrix.

Similarity gives an equivalence relation on the set of square matrices. The members of a class can be thought of as representing the same linear transformation but with respect to different bases for the associated space V. Canonical form theory then has the goal of finding "nice" matrices representing a given linear transformation—particularly matrices containing lots of zeros, with diagonal matrices as the grail. Such issues are a focus for Chapter 4.

3 Chapter 3: Determinants

The determinant is, at its most basic, a function from the set of all $n \times n$ matrices over $\mathbb{F}$ to $\mathbb{F}$ having properties that are important, particularly in the context of linear algebra. At a more general level it provides a concept of volume (area) that is not restricted by dimension or field of definition.

3.1 Introduction

The best known case of the determinant is that of 2×2 matrices:

$$\det \begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - bc.$$

Pleasant geometric arguments show that this is the area in the real plane of the parallelogram with adjacent sides given by the two vectors (a, b) and (c, d) (although your answer may turn out to be $bc - ad = -(ad - bc)$, depending upon the placement of the two vectors).

The geometry of 2×2 space and also the above formula confirm three properties:

(1) (i) The area of the parallelogram bounded by $r(a, b)$ and (c, d) is r times the area of the parallelogram bounded by (a, b) and (c, d).

(ii) The area of the parallelogram bounded by (a, b) and $(c + e, d + f)$ is the sum of the area of the parallelogram bounded by (a, b) and (c, d) and the area of the parallelogram bounded by (a, b) and (e, f).

(2) The area of the parallelogram bounded by (a, b) and any scalar multiple $\alpha(a, b)$ is 0.

(3) The parallelogram bounded by $(1, 0)$ and $(0, 1)$ is the unit square of area 1.

Area in dimension 2 corresponds to volume in dimension 3. In calculus and elsewhere the determinant of an $n \times n$ matrix is viewed as a (generalized, signed) volume, giving the volume of the n-parallelepiped bounded by a given set of n vectors at the origin. We will characterize the determinant as a generalized volume having properties extending the three above.

We have regularly considered the matrix A from $\mathrm{Mat}_{m,n}(\mathbb{F})$ as the n-tuple (system) $\mathcal{A} = (\mathbf{a}_1, \ldots, \mathbf{a}_j, \ldots, \mathbf{a}_n)$, where $\mathbf{a}_j$ is column j of A. In doing so, we are making an unspoken appeal to the natural vector space isomorphism

$$\eta \colon \mathrm{Mat}_{m,n}(\mathbb{F}) \longrightarrow (\mathbb{F}^m)_n \,,$$

this last space consisting of row n-tuples whose individual entries are column vectors from $\mathbb{F}^m$. For instance the correspondence

$$\eta\left(\begin{pmatrix} a & c \\ b & d \end{pmatrix}\right) = \left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix}\right)$$

played a role in our discussion of 2×2 determinants and planar area above.

In our discussion and development of determinant functions and determinants, we shall take both views of the $n \times n$ matrices under consideration. Specifically, we will consider determinant functions $D\colon (\mathbb{F}^n)_n \longrightarrow \mathbb{F}$, taking n-tuples of vectors from $\mathbb{F}^n$ to $\mathbb{F}$ and the related matrix functions $d\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$, the correspondence formally given by $D(\eta(A)) = d(A)$. For instance

$$\det \begin{pmatrix} a & c \\ b & d \end{pmatrix} = ad - bc = \mathrm{Det}\left(\begin{pmatrix} a \\ b \end{pmatrix}, \begin{pmatrix} c \\ d \end{pmatrix} \right) .$$

3.2 Properties determinant functions should have

In this section we consider "volume-like" functions taking n-tuples of vectors from $\mathbb{F}^n$ to $\mathbb{F}$. We call these determinant functions.

We consider determinant-like functions $D\colon (\mathbb{F}^n)_n \longrightarrow \mathbb{F}$ that satisfy the following natural generalizations of the first two "volume-like" properties of the previous section:

(I) **(n-Linear)** *Always* $D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, r\mathbf{u} + s\mathbf{w}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$
$= rD(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{u}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n) + sD(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{w}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$.

(II) **(Flat)** *Always* $D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) = 0$.

These can be restated:

(I) **(n-Linear)** *If we fix $n-1$ of the arguments of D, then D is linear in the remaining argument.*

(II) **(Flat)** *When two arguments of D are set equal, the value of D is* 0.

The following lemma describes the effect of elementary column operations on a determinant function. (See Proposition (2.3).) These properties also include those with which Treil §3.2 characterizes the determinant. In particular, Treil prefers the Antisymmetry Condition (3.1)(a) to our Flatness Condition (II). Provided the characteristic of $\mathbb{F}$ is not 2, Antisymmetry implies Flatness, since $-d = d$ if and only if $d = 0$. In the other direction, Flatness always Antisymmetry regardless of characteristic, as we see in the lemma.

(3.1). Lemma. *Let $D\colon (\mathbb{F}^n)_n \longrightarrow \mathbb{F}$ have* (I) *and* (II). *Then always*

(a) (Exchange: antisymmetry) *For $i \neq j$,*
$D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{u}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{w}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)$
$= -D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{w}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{u}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)$.

(b) (Scaling)
$D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, r\mathbf{u}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n) = rD(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{u}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$.

(c) (Column replacement) *For $j \neq k$,*

$D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{u} + s\mathbf{v}_j, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n) = D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{u}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n)$.

PROOF. (a) (II) with $\mathbf{v} = \mathbf{u} + \mathbf{w}$ then (I) four times and (II) twice again.
(b) (I) with $\mathbf{w} = \mathbf{0}$.
(c) By (b) and (II) $D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, s\mathbf{v}_j, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n) = 0$, so this follows from (I). □

The following easy consequences may be used without reference.

(3.2). COROLLARY. *Always*

(a) $D(\mathbf{v}_1, \ldots, \mathbf{v}_{k-1}, \mathbf{0}, \mathbf{v}_{k+1}, \ldots, \mathbf{v}_n) = 0$.

(b) $D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, r\mathbf{v}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) = 0$. □

(3.3). COROLLARY. *If the system* $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$ *is linearly dependent, then* $D(\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n) = 0$.

PROOF. This follows from the last part of the lemma and Proposition (1.5). □

The following technical result will also be of use.

(3.4). LEMMA. *Let D have properties* (I) *and* (II). *Then*

$$\begin{aligned} &D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &= (-1)^{i+j} D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_j, \mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &= (-1)^{i+j} D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_j, \mathbf{v}_i, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)\,. \end{aligned}$$

PROOF. Note that $(-1)^{j-i} = (-1)^{j-i}(-1)^{2i} = (-1)^{i+j}$. The proof is by induction on $k = j - i$, the number of steps from i to j. Lemma (3.1)(a) gives the initialization case $k = 1$ and also for $k > 1$

$$\begin{aligned} &D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &= -D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_j, \mathbf{v}_{j-1}, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)\,, \end{aligned}$$

and

$$\begin{aligned} &D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_i, \mathbf{v}_{i+1}, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n) \\ &= -D(\mathbf{v}_1, \ldots, \mathbf{v}_{i-1}, \mathbf{v}_{i+1}, \mathbf{v}_i, \ldots, \mathbf{v}_{j-1}, \mathbf{v}_j, \mathbf{v}_{j+1}, \ldots, \mathbf{v}_n)\,. \end{aligned}$$

The result then follows by induction. □

As seen in the proof, we can think of the result as saying:

> When you move one element of the system $[\![\mathbf{v}_1, \ldots, \mathbf{v}_j, \ldots, \mathbf{v}_n]\!]$ to another spot, the determinant changes by the factor $(-1)^k$, where k is the number of elements in the system that you have jumped.

3.3 Existence, uniqueness, and properties of determinants

The most familiar example of a determinant function is the matrix determinant $\det\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$, where, as before, we view a matrix as the list of its column vectors.

We shall say that a function $d\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$ has properties (I) and (II) if its associated vector function $D\,(= d\eta^{-1})$ has properties (I) and (II). The basic result is:

(3.5). THEOREM. *There is a unique function* det *with* (I) *and* (II) *and additionally having*

$$\det(I_n) = 1\,.$$

We will prove the theorem and, along the way, find various properties of the determinant and several ways of calculating it.

The proof of the theorem falls naturally into two parts—existence and uniqueness. There are (at least) four familiar ways to calculate a determinant:

- row expansion (row development);
- column expansion (column development);
- permutation sum;
- elementary column or row operations.

We prove existence using row expansion and uniqueness using elementary column operations. Later we discuss all the methods of calculation.

Authors typically use one of these methods of calculation as the initial definition of the determinant. Although TREIL develops the theory by using generalized volume, determinant functions, and elementary operations, his actual definition (in his Section 3.3.4) is via the permutation sum. We instead use row expansion on the first row as our initial definition, although ultimately we see that all are equivalent.

3.3.1 Existence of determinants

We show that determinant functions and determinants do exist. Uniqueness will then be proven in the next section.

If A is an $n\times n$ matrix, then $A^{(i,j)}$ is its $(n-1)\times(n-1)$ submatrix constructed by dropping row i and column j from A.

(3.6). PROPOSITION. (First row expansion) *The function* $\det_n\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *defined by* $\det_1(r) = r$ *(when* $n = 1$*) and for larger* n *iteratively by*

$$\det_n(A) = \sum_{j=1}^{n} (-1)^{1+j} a_{1j} \det_{n-1}(A^{(1,j)})\,.$$

has (I) *and* (II) *and* $\det_n(I_n) = 1$.

PROOF. The proof is by induction on n. The function $\det_1$ (with $\det_1(1) = 1$) certainly has (I), and (II) holds vacuously for $n = 1$. Now assume $n > 1$.

(I) Let $\mathbf{a}_k = r\mathbf{u} + s\mathbf{w}$ with $\mathbf{u} = \sum_{i=1}^n b_i\mathbf{e}_i$ and $\mathbf{w} = \sum_{i=1}^n c_i\mathbf{e}_i$. Set

$$B = [\![\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{u}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n]\!]$$

and

$$C = [\![\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, \mathbf{w}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n]\!] .$$

We prove that

$$\det_n(A) = r\det_n(B) + s\det_n(C) ,$$

by considering the individual terms for each j in the expansion (definition) of

$$\det_n(A) = \det_n(\mathbf{a}_1, \ldots, \mathbf{a}_{k-1}, r\mathbf{u} + s\mathbf{w}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n) .$$

If $j \neq k$, then

$$\det_{n-1}(A^{(1,j)}) = r\det_{n-1}(B^{(1,j)}) + s\det_{n-1}(C^{(1,j)})$$

by induction and property (I) for $\det_{n-1}$. Multiply throughout by the constant $(-1)^{1+j}a_{1j}$ to get the terms

$$(-1)^{1+j}a_{1j}\det_{n-1}(A^{(1,j)}) = r(-1)^{1+j}a_{1j}\det_{n-1}(B^{(1,j)}) + s(-1)^{1+j}a_{1j}\det_{n-1}(C^{(1,j)}) .$$

On the other hand for $j = k$ we have $a_{1k} = rb_1 + sc_1$ and $A^{(1,k)} = B^{(1,k)} = C^{(1,k)}$, so that also

$$(-1)^{1+k}a_{1k}\det_{n-1}(A^{(1,k)}) = r(-1)^{1+k}b_1\det_{n-1}(B^{(1,k)}) + s(-1)^{1+k}c_1\det_{n-1}(C^{(1,k)}) .$$

Therefore term-by-term

$$\det_n(A) = r\det_n(B) + s\det_n(C) ,$$

as desired and giving (I).

(II) We must examine

$$\det_n(A) = \det_n(\mathbf{a}_1, \ldots, \mathbf{a}_{i-1}, \mathbf{a}, \mathbf{a}_{i+1}, \ldots, \mathbf{a}_{k-1}, \mathbf{a}, \mathbf{a}_{k+1}, \ldots, \mathbf{a}_n)$$

with $\mathbf{a}_i = \mathbf{a}_k = \mathbf{a}$. Again we consider the individual terms for each j.

If $j \notin \{i, k\}$, then $A^{(1,j)}$ contains the repeated column $\mathbf{a}_i^{(1)} = \mathbf{a}^{(1)} = \mathbf{a}_k^{(1)}$ (the various columns with their first entries deleted). Thus for these j we have $\det_{n-1}(A^{(1,j)}) = 0$ by (II) for $\det_{n-1}$.

For $j \in \{i, k\}$, the submatrices $A^{(1,i)}$ and $A^{(1,k)}$ are the same except that the vector $\mathbf{a}^{(1)}$ is in different positions—the first contains the column subsequence

$$\ldots, \mathbf{a}_{i-1}^{(1)}, \mathbf{a}_{i+1}^{(1)}, \ldots, \mathbf{a}_{k-1}^{(1)}, \mathbf{a}^{(1)}, \mathbf{a}_{k+1}^{(1)} \ldots$$

and the second

$$\ldots, \mathbf{a}_{i-1}^{(1)}, \mathbf{a}^{(1)}, \mathbf{a}_{i+1}^{(1)}, \ldots, \mathbf{a}_{k-1}^{(1)}, \mathbf{a}_{k+1}^{(1)} \ldots .$$

From Lemma (3.4) we find

$$\det_{n-1}(A^{(1,i)}) = (-1)^{i+k-1}\det_{n-1}(A^{(1,k)})\,.$$

Also since $\mathbf{a}_i = \mathbf{a}_k = \mathbf{a}$ we have $a_{1i} = a_{1k}$. Combining all the terms, we have

$$\begin{aligned}\det_n(A) &= \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det_{n-1}(A^{(1,j)}) \\ &= (-1)^{1+i}a_{1i}\det_{n-1}(A^{(1,i)}) + (-1)^{1+k}a_{1k}\det_{n-1}(A^{(1,k)}) \\ &= a_{1k}\left((-1)^{1+i}(-1)^{i+k-1}\det_{n-1}(A^{(1,k)}) + (-1)^{1+k}\det_{n-1}(A^{(1,k)})\right) \\ &= 0\,.\end{aligned}$$

This gives (II).

Finally for $A = I_n$ we have $a_{11} = 1$ and $a_{1j} = 0$ for $j > 1$. Also $A^{(1,1)} = I_{n-1}$, so

$$\begin{aligned}\det_n(I_n) &= \sum_{j=1}^{n}(-1)^{1+j}a_{1j}\det_{n-1}(A^{(1,j)}) = (-1)^{1+1}\cdot 1\cdot\det_{n-1}(A^{(1,1)}) \\ &= 1\cdot 1\cdot\det_{n-1}(I_{n-1}) = 1\,,\end{aligned}$$

as claimed. □

3.3.2 Uniqueness of determinants

As an immediate consequence of Gaussian elimination, we find in Theorem (2.4) that every matrix is the product of elementary matrices and a matrix in echelon form. In particular, every square matrix is the product of elementary matrices followed by an upper triangular matrix (0's below the diagonal). By transposing, we then have that every square matrix can be written as the product of a lower triangular matrix followed by a product of elementary matrices.

(3.7). THEOREM. *Let* $\det\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *be a function satsifying* (I) *and* (II) *and additionally having* $\det(I_n) = 1$.

(a) $\det(X_{i,j}) = -1$*;* $\det(S_i(r)) = r$*;* $\det(R_{i,j}(s)) = 1$.

(b) *If* $T \in \mathrm{Mat}_{n,n}(\mathbb{F})$ *is a triangular matrix, then* $\det(T)$ *is the product of the diagonal entries of* T.

(c) *For* $A \in \mathrm{Mat}_{n,n}(\mathbb{F})$*, let* $A = T\prod_{i=1}^{k} E_i$ *where* T *is a triangular matrix and the* E_i *are elementary matrices. Then* $\det(A) = \det(T)\prod_{i=1}^{k}\det(E_i)$.

Before discussing the proof of this we observe an immediate corollary.

(3.8). COROLLARY. *For every* n *there is at most one function* $\det\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *satsifying* (I) *and* (II) *and additionally having* $\det(I_n) = 1$.

PROOF. Indeed, using the theorem and the remarks that precede it, we can calculate all the values of such a function. □

This is uniqueness for the determinant. What is not clear from this result is that such a function exists. It is conceivable that different factorizations $A = T\prod_{i=1}^{k} E_i = T'\prod_{i=1}^{k} E_i'$ give conflicting values $\det(T)\prod_{i=1}^{k}\det(E_i)$ and $\det(T')\prod_{i=1}^{k}\det(E_i')$, in which case the function would not be well-defined and so cannot exist. Luckily Proposition (3.6) prevents this:

PROOF OF THEOREM (3.5).
A determinant map $\det = \det_n$ exists for all n by Proposition (3.6) and is unique by Corollary (3.8). □

Having discussed these important consequences, we now proceed to the proof of Theorem (3.7).

Throughout the remainder of Section 3.3, we consider a function

$$\det\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$$

that has (I) and (II) and additionally $\det(I_n) = 1$. We shall at times refer to this function as "the determinant" although we do not yet know that it is unique.

3.3.3 Diagonal and triangular matrices

See Proposition (3.10) of the next section.

3.3.4 Calculation of determinants using elementary operations

(3.9). LEMMA. *Let A be an $n \times n$ matrix over $\mathbb{F}$.*

(a) (Exchange) *For $i \neq j$,* $\det(AX_{i,j}) = -\det(A)$.

(b) (Scaling) $\det(AS_k(r)) = r\,\det(A)$, *for* $0 \neq r \in \mathbb{F}$.

(c) (Column replacement) *For $j \neq k$,* $\det(AR_{j,k}(s)) = \det(A)$.

PROOF. This is Lemma (3.1) rewritten in matrix terms. □

(3.10). PROPOSITION. *The determinant of a triangular matrix is the product of its diagonal entries.*

PROOF. If none of the diagonal entries are 0, then multiplication by column replacement matrices as in Lemma (3.9)(c) leave the determinant unchanged while moving to a diagonal matrix with the same nonzero diagonal entries. The determinant of this diagonal matrix is the product of its diagonal entries by scaling as in Lemma (3.9)(b).

If there are 0's on the diagonal, then the first diagonal 0 is in a nonpivot column of the eventual echelon form, so the matrix is not of full rank. But then its determinant is 0 by Corollary (3.3). □

(3.11). PROPOSITION. *Let* $\det\colon \mathrm{Mat}_{n,n}(\mathbb{F}) \longrightarrow \mathbb{F}$ *be a function satsifying* (I) *and* (II) *and additionally having* $\det(I_n) = 1$.

(a) $\det(X_{i,j}) = -1$; $\det(S_i(r)) = r$; $\det(R_{i,j}(s)) = 1$.

(b) *For* $A \in \mathrm{Mat}_{n,n}(\mathbb{F})$, *let* $A = B\prod_{i=1}^{k} E_i$ *where* $B \in \mathrm{Mat}_{n,n}(\mathbb{F})$ *and the* E_i *are elementary matrices. Then* $\det(A) = \det(B)\prod_{i=1}^{k}\det(E_i)$.

PROOF. The first part comes from Lemma (3.9) with $A = I_n$. The same lemma and (a) then give the case $k = 1$ of (b) at which point the rest follows by induction on k. □

PROOF OF THEOREM (3.7).
This is immediate from Propositions (3.10) and (3.11). □

3.3.5 Some nice determinant properties

(3.12). PROPOSITION. *A square matrix is invertible if and only if its determinant is nonzero.*

PROOF. From Theorem (2.4) we get $A = T\prod_{i=1}^{k} E_i$, where T is a triangular matrix and the E_i are elementary matrices. By Theorem (3.7) the matrix A is invertible if and only if T is invertible, and the proposition follows from Proposition (3.10). □

(3.13). THEOREM. *Let* A *and* B *be square matrices with entries from the field* $\mathbb{F}$. *Then* $\det(AB) = \det(A)\det(B)$.

PROOF. If A or B is not invertible, then AB is also not invertible by Corollary (2.18). In that case both sides of the equality are 0 by the previous proposition.

We may now assume that A and B are invertible. By Corollary (2.16) we have $A = \prod_{i=1}^{k} E_i$ and $B = \prod_{j=1}^{l} F_j$ for elementary matrices E_i and F_j. Of course $AB = \prod_{i=1}^{k} E_i \prod_{j=1}^{l} F_j$, so the equality follows from the last part of Theorem (3.7). □

The previous theorem gives one of the most important properties of the determinant. Indeed there are places where this property, together with the requirement (as in Proposition (3.10)) that a diagonal matrix have the product of its diagonal entries as determinant, is used as the definition of the determinant.

(3.14). THEOREM. *Let* A *be a square matrix with entries from the field* $\mathbb{F}$. $\det(A) = \det(A^\top)$.

PROOF. By Proposition (3.12) we need only consider matrices A that are invertible. By Theorem (2.16) $A = \prod_{i=1}^{m} E_i$ as a product of elementary matrices. Then $A^\top = \prod_{i=m}^{1} E_i^\top$. Every elementary matrix is either triangular or equal to its own transpose, hence $\det E_i = \det E_i^\top$ (by Proposition (3.10)). Thus this theorem follows from the previous one. □

3.3.6 Summary of properties of determinants

(1) The determinant is linear in each column (resp., row) when the other columns (resp., rows) are fixed: (I) and Theorem (3.14).

(2) The exchange of two columns (resp., rows) negates the determinant: Lemma (3.9)(a) and Theorem (3.14).

(3) The determinant of a triangular matrix is the product of the diagonal entries: Proposition (3.10).

(4) If a matrix has a zero column or row, then it has determinant 0: Corollary (3.2)(a) and Theorem (3.14).

(5) If a matrix has two equal columns or rows, then it has determinant 0: (II) and Theorem (3.14).

(6) If one of the columns (resp., rows) of a matrix is a linear combination of the remaining columns (resp., rows), then the matrix has determinant 0: Corollary (3.3) and Theorem (3.14).

(7) $\det(A) = 0$ if and only if A is not invertible: Proposition (3.12).

(8) $\det(A) \neq 0$ if and only if A is invertible: Proposition (3.12).

(9) The determinant is not changed by column (resp., row) replacement: Lemma (3.9)(c) and Theorem (3.14).

(10) $\det(A) = \det(A^\top)$: Theorem (3.14).

(11) $\det(AB) = \det(A)\det(B)$: Theorem (3.13).

(12) If A is $n \times n$, then $\det(aA) = a^n \det(A)$: Lemma (3.9)(b) (n times).

3.4 Permutation sum expansion

Let $\mathcal{E} = [\![\mathbf{e}, \ldots, \mathbf{e}_n]\!]$ be the standard basis of $\mathbb{F}^n$. For $A = (a_{i,j})_{i,j} \in \mathrm{Mat}_{n,n}(\mathbb{F})$ we have

$$\begin{aligned}\det(A) &= \det(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{a}_j, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n) \\ &= \sum_{i=1}^{n} a_{ij} \det(\mathbf{a}_1, \ldots, \mathbf{a}_{j-1}, \mathbf{e}_i, \mathbf{a}_{j+1}, \ldots, \mathbf{a}_n)\end{aligned}$$

by n-linearity (I). If instead we write every $\mathbf{a}_j$ as a linear combination of basis elements from $\mathcal{E}$ and then expand in every coordinate, we arrive at

$$\det(A) = \sum_{i_1=1}^{n} \sum_{i_2=1}^{n} \cdots \sum_{i_n=1}^{n} a_{i_1,1} a_{i_2,2} \cdots a_{i_n,n} \det(\mathbf{e}_{i_1}, \mathbf{e}_{i_2}, \ldots, \mathbf{e}_{i_n}),$$

which can be better written as

$$\det(A) = \sum_{\varphi \in \mathrm{Fun}(n)} \left(\prod_{j=1}^{n} a_{\varphi(j),j} \right) \det(\mathbf{e}_{\varphi(1)}, \mathbf{e}_{\varphi(2)}, \ldots, \mathbf{e}_{\varphi(n)}) \,,$$

where $\mathrm{Fun}(n)$ is the set of all functions from $\{1, \ldots, n\}$ to $\{1, \ldots, n\}$. Calculation of the determinant is thus reduced to finding

$$\det(\mathbf{e}_{\varphi(1)}, \mathbf{e}_{\varphi(2)}, \ldots, \mathbf{e}_{\varphi(n)})$$

for the various $\varphi \in \mathrm{Fun}(n)$.

Many of these are easy to calculate. If different a and b have $\varphi(a) = \varphi(b)$, then $[\![\mathbf{e}_{\varphi(1)}, \mathbf{e}_{\varphi(2)}, \ldots, \mathbf{e}_{\varphi(n)}]\!]$ contains a repeat and $\det(\mathbf{e}_{\varphi(1)}, \mathbf{e}_{\varphi(2)}, \ldots, \mathbf{e}_{\varphi(n)}) = 0$ by (II). Therefore we need only sum over the set of functions that are bijections of $\{1, \ldots, n\}$ with itself. This is the set $\mathrm{Perm}(n)$ of all *permutations* of $\{1, \ldots, n\}$.

For a permutation σ, a sequence of "column exchange" operations turns the list $[\![\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)}]\!]$ into $[\![\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]\!]$, the $n \times n$ identity matrix. Therefore by Lemma (3.1)(a)

$$\det(\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)}) = (-1)^k \det(\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n) = (-1)^k \cdot 1 = (-1)^k \,,$$

where k is the number of exchanges made.

The number $(-1)^k$ is called the *sign* (or *signum*) of the permutation σ and is denoted $\mathrm{sgn}(\sigma)$. There may be many different sequences of exchanges that move us from $[\![\mathbf{e}_{\sigma(1)}, \mathbf{e}_{\sigma(2)}, \ldots, \mathbf{e}_{\sigma(n)}]\!]$ to $[\![\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_n]\!]$, but they all must have the same sign by Theorem (3.5) and the above. [6] That is, if one sequence for σ contains an odd number k of exchanges, then all sequences for σ have an odd number of exchanges (although not necessarily k). In this case σ is called an *odd* permutation. On the other hand, if one sequence for σ uses an even number of exchanges then they all do, and σ is an *even* permutation.

This discussion proves

(3.15). THEOREM. (Permutation sum expansion) *Let $A = (a_{ij})_{ij}$ be an $n \times n$ matrix with entries from the field $\mathbb{F}$. Then*

$$\det(A) = \sum_{\sigma \in \mathrm{Perm}(n)} \mathrm{sgn}(\sigma) \prod_{j=1}^{n} a_{\sigma(j),j} \,. \qquad \square$$

3.5 Column and row expansion

(3.16). THEOREM. (Column expansion) *Let $A = (a_{ij})_{ij}$ be an $n \times n$ matrix with entries from the field $\mathbb{F}$. Then for every column index j*

$$\det(A) = \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A^{(i,j)}) \,.$$

[6] Note that since TREIL uses the permutation sum to demonstrate formal existence of the determinant, he must (unlike us) independently prove that sgn is well-defined. He discusses this on page 88.

PROOF. For column $j = 1$ this follows from first row expansion (Proposition (3.6)) and the invariance of determinant under transpose (Theorem (3.14)).

Let B be the matrix A with column j of A moved to column 1 of B with the remaining columns sliding to the right. Then always $a_{i,j} = b_{i,1}$ and $A^{(i,j)} = B^{(i,1)}$.

Thereofore by Lemma (3.4) and the column 1 case,

$$\begin{aligned}\det(A) &= (-1)^{j+1} \det(B) \\ &= (-1)^{j+1} \sum_{i=1}^{n} (-1)^{i+1} b_{i1} \det(B^{(i,1)}) \\ &= \sum_{i=1}^{n} (-1)^{i+j} a_{ij} \det(A^{(i,j)}) \,,\end{aligned}$$

as desired. □

(3.17). THEOREM. (Row expansion) *Let $A = (a_{ij})_{ij}$ be an $n \times n$ matrix with entries from the field $\mathbb{F}$. Then for every row index i*

$$\det(A) = \sum_{j=1}^{n} (-1)^{i+j} a_{ij} \det(A^{(i,j)}) \,.$$

PROOF. This is an immediate consequence of the previous theorem, using the transpose and Theorem (3.14). □

3.6 Minors

This material was not covered in the course.

4 Chapter 4: Introduction to spectral theory

Chapter 2 and Gaussian elimination were motivated by the solving of linear equations. We wish for representatives and canonical forms for row equivalence—especially, we want representatives containing lots of zeros.

At the end of that chapter we introduced matrix similarity as the equivalence relation associated with choosing different bases for representation of a fixed linear transformation. A large part of this chapter is devoted to finding representatives and canonical forms for similarity; again we prize representatives that contain many zeros.

A square matrix is *triangulable* or *triangularizable* if it is similar to a triangular matrix. In the special case when it is similar to a diagonal matrix it may also be called *diagonable* or *diagonalizable* (this is the most frequent terminology and is preferred by Treil). Although these are critical concepts in this chapter, we avoid the specific terminology.

Parts of this chapter can be made valid for infinite dimensional spaces (when stated appropriately), but we shall only consider finite dimensional spaces.

4.1 Main definitions

4.1.1 Eigenvalue, eigenvectors, and spectrum

If T is a linear transformation of V (that is, from V to V), then the nonzero vector $\mathbf{v}$ is an *eigenvector* associated with the *eigenvalue* $\lambda \in \mathbb{F}$ when

$$T(\mathbf{v}) = \lambda \mathbf{v}\,.$$

The *spectrum* of T is then the set $\sigma(T)$ of all its eigenvalues.

Of particular interest are the eigenvalues and eigenvectors of the $n \times n$ matrix A from $\mathbb{F}$, viewed as the matrix linear transformation $[A]\colon \mathbb{F}^n \longrightarrow \mathbb{F}^n$ given by $\mathbf{v} \mapsto A\mathbf{v}$. It spectrum is $\sigma(A) = \sigma([A])$.

The motivation here is that the standard basis of $\mathbb{F}^n$ is a basis of eigenvectors for a diagonal matrix from $\mathrm{Mat}_{n,n}(\mathbb{F})$, its diagonal elements being the corresponding eigenvalues.

(4.1). Theorem. *Let T be a linear transformation of the $\mathbb{F}$-space V of dimension n, and let $\lambda \in \mathbb{F}$. The following are equivalent:*

(1) *λ is an eigenvalue of T;*

(2) $\mathrm{Ker}(T - \lambda I)$ *is nonzero;*

(3) *there is a basis $\mathcal{B}$ of V with* $[T]_{\mathcal{B}\mathcal{B}} = \begin{pmatrix} \lambda & \mathbf{a} \\ \mathbf{0} & B \end{pmatrix}$, *where $\mathbf{0}$ is the zero vector of $\mathbb{F}^{n-1}$, $\mathbf{a}$ is some vector of $\mathbb{F}_{n-1} = \mathrm{Mat}_{1,n-1}(\mathbb{F})$, and B is some matrix of* $\mathrm{Mat}_{n-1,n-1}(\mathbb{F})$.

PROOF. For $\mathbf{v} \neq \mathbf{0}$ and $\lambda \in \mathbb{F}$

$$\mathbf{v} \in \mathrm{Ker}(T - \lambda I) \iff \mathbf{0} = (T - \lambda I)(\mathbf{v}) = T(\mathbf{v}) - \lambda \mathbf{v} \iff T(\mathbf{v}) = \lambda \mathbf{v}\,.$$

Therefore (1) and (2) are equivalent.

Always $[T]_{\mathcal{B}\mathcal{B}}[\mathbf{v}]_{\mathcal{B}} = [T(\mathbf{v})]_{\mathcal{B}}$ for the basis $\mathcal{B} = [\![\mathbf{b}_1, \dots]\!]$. In particular $[T]_{\mathcal{B}\mathcal{B}} = \begin{pmatrix} \lambda & \mathbf{a} \\ \mathbf{0} & B \end{pmatrix}$ if and only if $(1, 0, \dots, 0)^\top = [\mathbf{b}_1]_{\mathcal{B}}$ is an eigenvector of $[T]_{\mathcal{B}\mathcal{B}}$ for the eigenvector λ, which is in turn true if and only if $\mathbf{b}_1$ is an eigenvector of T for the eigenvector λ. Thus (1) holds if and only if (3) holds. □

4.1.2 Eigenvalues and the characteristic polynomial

See Section 4.1.4 below.

4.1.3 Characteristic polynomial of an operator

See Section 4.1.4 below.

4.1.4 Multiplicities and the characteristic polynomial

Theorem (4.1) makes it clear that the set of all eigenvectors associated with a fixed eigenvalue λ is a subspace of V (indeed, the kernel of $T - \lambda I$). This is the *eigenspace* of λ. Here and elsewhere we abuse our terminology somewhat: by definition all eigenvectors are nonzero; nevertheless the zero vector $\mathbf{0}$ belongs to the eigenspace for each eigenvalue λ. (Think of $\mathbf{0}$ as a *weak* eigenvector.) For eigenvalue λ, the dimension of its eigenspace is its *geometric multiplicity*.

(4.2). PROPOSITION. *Let T be a linear transformation of the $\mathbb{F}$-space V of dimension n, and let $\lambda \in \mathbb{F}$. For the positive integer h, the following are equivalent:*

(1) *λ is an eigenvalue of T of geometric multiplicity at least h;*

(2) *there is a basis $\mathcal{B}$ of V with*

$$[T]_{\mathcal{B}\mathcal{B}} = \begin{pmatrix} \lambda I_h & C \\ 0 & D \end{pmatrix}\,.$$

PROOF. As in Theorem (4.1), the basis $\mathcal{B} = [\![\mathbf{b}_1, \dots, \mathbf{b}_h, \mathbf{b}_{h+1}, \dots]\!]$ includes $\mathbf{b}_1, \dots, \mathbf{b}_h$ spanning a subspace consisting of (weak) eigenvectors of T for λ if and only if $[T]_{\mathcal{B}\mathcal{B}}$ has the given form. □

By Corollary (3.3) and Theorem (4.1) we know that λ is an eigenvalue of the matrix A if and only if it is a zero of the function $\det(A - zI)$. This function is very important.

(4.3). LEMMA. *If A is an $n \times n$ matrix from $\mathbb{F}$, then $\det(A - zI)$ is a polynomial function of degree n in the variable z.*

PROOF. This follows from permutation sum expansion of the determinant, Theorem (3.15) (or from row or column expansion and induction). □

We call $\det(A - zI)$ the *characteristic polynomial* of a matrix A and denote it by $\mathrm{cp}_A(z)$.

(4.4). LEMMA. *Similar matrices have the same characteristic polynomial.*

PROOF. This follows from Theorem (3.13). □

In particular we can define the *characteristic polynomial* $\mathrm{cp}_T(z)$ of the linear transformation (operator) T as the characteristic polynomial of any matrix A that represents it. Since every square matrix represents some linear transformation, we usually restrict our discussion to characteristic polynomials of square matrices, always remembering that each result could be easily restated in terms of linear transformations.

We have an extension of Theorem (4.1):

(4.5). THEOREM. *Let A be an $n \times n$ matrix from $\mathbb{F}$, and let $\lambda \in \mathbb{F}$. The following are equivalent:*

(1) *λ is an eigenvalue of A;*

(2) *$\mathrm{NS}(A - \lambda I)$ is nonzero;*

(3) *A is similar to a matrix $\begin{pmatrix} \lambda & \mathbf{a} \\ \mathbf{0} & B \end{pmatrix}$ for $\mathbf{0} \in \mathbb{F}^{n-1}$, $\mathbf{a} \in \mathrm{Mat}_{1,n-1}(\mathbb{F})$, $B \in \mathrm{Mat}_{n-1,n-1}(\mathbb{F})$;*

(4) *$\det(A - \lambda I) = 0$;*

(5) *λ is a root of the characteristic polynomial $\mathrm{cp}_A(z) = \det(A - zI)$;*

(6) *$\mathrm{cp}_A(z) = (\lambda - z)q(z)$ for a polynomial function $q(z)$ of degree $n - 1$.*

PROOF. The first three are equivalent by Theorem (4.1).

(4) is equivalent to (2) by Proposition (3.12) and to (5) by the definition of the characteristic polynomial. Finally (5) and (6) are true by basic properties of polynomials. □

By Proposition (4.2) and Lemma (4.4)

$$\mathrm{cp}_A(z) = (\lambda - z)^g p(z)$$

where g is the geometric multiplicity of λ as an eigenvalue of A and $p(z)$ is some polynomial of degree $n - g$. The *algebraic multiplicity* of the eigenvalue λ of A is the largest integer a with

$$\mathrm{cp}_A(z) = (\lambda - z)^a o(z)$$

for some polynomial $o(z)$ of degree $n - a$. Clearly then

(4.6). Proposition. (Treil's Proposition 4.1.1.) *If λ is an eigenvalue of A with geometric multiplicity g and algebraic multiplicity a then $1 \leq g \leq a$.* □

(4.7). Proposition. *Let A be an $n \times n$ matrix from $\mathbb{F}$, and let $\lambda \in \mathbb{F}$. For the positive integer b, the following are equivalent:*

(1) *λ is an eigenvalue of A of algebraic multiplicity at least b;*

(2) *A is similar to a matrix*

$$\begin{pmatrix} \lambda & * & * & \\ 0 & \ddots & * & C^{b,n-b} \\ 0 & 0 & \lambda & \\ & \mathbf{0}^{n-b,b} & & D^{n-b,n-b} \end{pmatrix}.$$

Proof. That (2) implies (1) is clear.

Assuming (1), we have similarity to a matrix $\begin{pmatrix} \lambda & \mathbf{a} \\ \mathbf{0}^{n-1,1} & B \end{pmatrix}$ by Theorem (4.1). Here λ is an eigenvalue of B of algebraic multiplicity at least $b-1$, and we are done by induction. □

(4.8). Lemma. *If $\mathbf{v}$ is an eigenvector of A associated with the eigenvalue λ and $p(x)$ is a polynomial, then $\mathbf{v}$ is also an eigenvector of $p(A)$, now associated with the eigenvalue $p(\lambda)$. The geometric and algebraic multiplicities of $p(\lambda)$ as an eigenvalue of $p(A)$ are at least equal to those of λ as an eigenvalue of A.*

Proof. This is a direct consequence of Lemma (4.4) and Propositions (4.2) and (4.7). □

4.1.5 Trace and determinant

See Corollary (4.11) below.

4.1.6 Similarity and triangular matrices

(4.9). Theorem. *The $n \times n$ matrix A with entries from $\mathbb{F}$ is similar to a triangular matrix with diagonal entries $\lambda_1, \ldots, \lambda_n$ if and only if*

$$\mathrm{cp}_A(z) = \prod_{i=1}^{n} (\lambda_i - z).$$

Proof. Use Proposition (4.7) and induction. □

By the Fundamental Theorem of Algebra, we get immediately:

(4.10). Corollary. *Every complex $n \times n$ matrix is similar to a triangular matrix.* □

(4.11). Corollary. (Treil's Theorem 4.1.2.) *Let A be an $n \times n$ matrix with eigenvalues $\lambda_1, \ldots, \lambda_n$ (including algebraic multiplicities). Then*

(a) $\det A = \prod_{i=1}^{n} \lambda_i$;

(b) $\operatorname{tr} A = \sum_{i=1}^{n} \lambda_i$.

PROOF. By Lemma (4.4) and Theorem (4.9) we need only consider triangular matrices, for which the result is clearly true. □

4.2 Similarity and diagonal matrices

(4.12). LEMMA. (TREIL's Theorem 4.2.2.) *If*

$$\lambda_1, \ldots, \lambda_r$$

are distinct eigenvalues of T with associated eigenvectors

$$\mathbf{v}_1, \ldots, \mathbf{v}_r$$

then the system $[\![\mathbf{v}_1, \ldots, \mathbf{v}_r]\!]$ is linearly independent.

PROOF. Pages 106–7 of TREIL presents a nice induction proof of this. □

(4.13). THEOREM. (See TREIL's Theorems 4.2.1 and 4.2.8.) *Let A be an $n \times n$ matrix over $\mathbb{F}$ that is similar to a triangular matrix. Then the following are equivalent:*

(1) *A is similar to a diagonal matrix;*

(2) *$\mathbb{F}^n$ possesses a basis of eigenvectors for A;*

(3) *the geometric multiplicity of each eigenvalue of A is equal to its algebraic multiplicity.*

PROOF. For the diagonal matrix D with $\lambda_1, \ldots, \lambda_n$ down the diagonal

$$Q^{-1}AQ = D \iff AQ = QD\,,$$

and we see that column j of Q is an eigenvector of A for the eigenvalue λ_j. Therefore (1) implies (2).

Conversely, if $\mathcal{B}$ is a basis of eigenvectors for A and Q is the matrix whose columns are the members of Q then $A = [A]_{\mathcal{E}\mathcal{E}}$ and $Q = [I]_{\mathcal{E}\mathcal{B}}$ gives

$$Q^{-1}AQ = [I]_{\mathcal{B}\mathcal{E}}[A]_{\mathcal{E}\mathcal{E}}[I]_{\mathcal{E}\mathcal{B}} = [A]_{\mathcal{B}\mathcal{B}} = D\,,$$

a diagonal matrix. Thus (2) implies (1).

Clearly (1) and (2) imply (3).

We conclude by proving that (3) implies (2). Assume $\operatorname{cp}_A(z) = \prod_{i=1}^{t}(\lambda_i - z)^{a_i}$, where a_i is the algebraic and geometric multiplicity of the eigenvalue λ_i. For each i let $\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,a_i}$ be a basis of the eigenspace V_i for λ_i. We claim that the system

$$\mathcal{B} = [\![\mathbf{v}_{1,1}, \ldots, \mathbf{v}_{1,a_1}, \ldots, \mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,a_i}, \ldots, \mathbf{v}_{t,1}, \ldots, \mathbf{v}_{t,a_t}]\!]$$

is a basis of eigenvectors for A on $\mathbb{F}^n$. It certainly consists of eigenvectors, and its size is $\sum_{i=1}^{t} a_i = n$; so we only need to prove that it is linearly independent. Suppose

$$\sum_{i=1}^{t} \sum_{j=1}^{a_i} \alpha_{i,j} \mathbf{v}_{i,j} = \mathbf{0}\,.$$

That is, $\sum_{i=1}^{t} \mathbf{v}_i = \mathbf{0}$ for $\mathbf{v}_i = \sum_{j=1}^{a_i} \alpha_{i,j} \mathbf{v}_{i,j} \in V_i$. The various $\mathbf{v}_i$ are (weak) eigenvectors for the distinct eigenvalues λ_i, so by Lemma (4.12) each is $\mathbf{0}$. But then as the corresponding $[\![\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,a_i}]\!]$ is a basis of V_i, we find that every $\alpha_{i,j}$ is 0. $\mathcal{B}$ is indeed a basis. □

(4.14). Corollary. (Treil's Theorem 4.2.3.) *If the $n \times n$ matrix A has n distinct eigenvalues, then it is similar to a diagonal matrix.*

Proof. This follows from the theorem and Lemma (4.12). □