Math 4242 Sec 40

Supplemental Notes + Homework 8 (With Solutions)

1 Infinite Bases and Dimension

Though in this course we care mostly about vector spaces with finite bases, even for this it is convenient to have the language of infinite bases at hand. With that in mind, I'll present the definitions of span, linearly independence, and basis, in a way that make sense in both the finite and infinite settings. The definitions are small tweaks of what I already presented in class.

For S a (possibly infinite) set in a vector space V over a field F, a **linear combination** of elements in S is a **finite** sum

$$c_1 s_1 + c_2 s_2 + \cdots + c_k s_k,$$

with each $c_i \in F$ and each $s_i \in S$.

As before, we define $\text{Span}(S)$, the **span** of S, to be the set of all linear combinations of elements in S. $\text{Span}(S)$ is a subspace of V.

A (possibly) infinite set $S \in V$ is **linearly independent** if whenever $c_1 s_1 + c_2 s_2 + \cdots + c_k s_k = 0$ for some $c_i \in F$ and the s_i distinct elements of S, we have that each $c_i = 0$. We regard an empty set of vectors as linearly independent.

For V a vector space, a (possibly) infinite set B is a **basis** for V if

1. $\text{Span}(B) = V$,
2. B is linearly independent.

Note: the empty set is a basis for a "trivial" vector space containing only single element $\vec{0}$.

For a set S, we define $|S|$ to be the number of elements in S, if S is finite. Otherwise, we define $|S| = \infty$.

Now we can state the result we proved (most of) in friday's class, in a more general form that makes sense for both finite and infinite bases:

Proposition 1.1. *If B and D are both bases for a vector space V, then $|B| = |D|$.*

Note that in particular Proposition 1.1 tells us that if a vector space V has a finite basis, then all bases for V are finite.

Proof of Proposition 1.1. To keep notation simple, we'll restrict attention to the special case that B and D are finite; essentially the same proof gives the result in general, though the notation gets a little more complex.

I'll give the same proof from class, with some easy details omitted.

Write $B = \{b_1, \ldots, b_k\}$. For each $0 \leq q \leq k$, we construct bases S_q for V by removing q elements of D and adding in the first q elements of Q. We do this inductively, first constructing S_1, by removing a single element of D and adding in b_1; then we construct S_2 by removing a second element of D from S_1 and adding in b_2; and so on, until we have constructed S_k. Then in particular, S_k is a basis for V containing B. But since a basis for V is a minimal spanning set for V, we must have that $S_k = B$. Since $|S_k| = |D|$ by construction, we have $|B| = |D|$, which gives the result.

For the case $q = 0$, we take $S_0 = B$. This is clearly a basis. Now assume that for some q, $1 \leq q \leq k$, we have constructed S_{q-1} as above. We construct S_q. Write $S_{q-1} = \{s_1, \ldots s_l\}$. Since S_{q-1} is a basis, $b_q \in \mathrm{Span}(S_{q-1})$, so we may write

$$b_q = \sum_{j=1}^{l} c_j s_j$$

for some scalars c_j. Since B is linear independent, so is any subset of B (make sure you understand why). Then, by the linear independence of $\{b_1, \ldots, b_q\}$, there must exist some $s_i \in S_{q-1} \cap D$ such that c_i is non-zero.

We define $S_q = S_q \cup \{b_q\} - \{s_i\}$; in other words, we form S_q from S_{q-1} by adding in b_q and removing s_i. Given the assumptions on S_{q-1} it is clear that S_q is formed by removing a set D_q of size q from D and adding in the elements $\{b_1, \ldots, b_q\}$, as desired.

It remains to check that S_q is really a basis. We verified this in class for the case $q = 1$. The verification for arbitrary q is the same and is straightforward. I'll omit it here. $\square$

Proposition 1.2. *Any vector space has a basis.*

We'll omit the proof of Proposition 1.2. The proof, while not especially complicated, requires the **axiom of choice**, an axiom from set theory, that I'd rather not discuss now.

Having presented the above results and definitions, we can now define the dimension of an arbitrary vector space V by taking $\dim(V) = |B|$, for B any basis of V. We say a vector space V is **finite-dimensional** if $\dim(V) < \infty$.

Example 1.3.

1. $\dim(\mathbb{R}^n) = n$; the standard basis for $\mathbb{R}^n$ has n elements.

2. For any $c \in \mathbb{R}$, $V = \{(x, y) \in \mathbb{R}^n \mid y = cx\}$ is a vector space with $\dim(V) = 1$.

3. For any vector space V, $\{\vec{0}\}$ is a subspace of V of dimension 0; the empty set is a basis.

4. The vector space P of all polynomial functions $\mathbb{R} \to \mathbb{R}$ is an example of an infinite-dimensional vector space;

$$S = \{1, x, x^2, \ldots\}$$

is a basis for P. It's easy to see that $\mathrm{Span}(S) = P$. Linear independence of S follows from the fact that a non-zero polynomial of degree n has at most n roots.

The following technical proposition yields a couple of intuitive and useful results about subspaces:

Proposition 1.4. *If V is a finite-dimensional vector space and W is a subspace of V, then any basis for W can be extended to a basis for V.*

Proof. Let B_0 be a basis for W. Choose a basis $D = \{v_1, \cdots, v_k\}$ for V.

Now for $1 \le i \le l$, define B_i inductively, by taking $B_i = B_{i-1}$ if $v_i \in \mathrm{Span}(B_{i-1})$ and $B_i = B_{i-1} \cup \{v_i\}$ otherwise. Note that $B_0 \subseteq B_k$. We'll show that B_k is a basis for V, which gives the result.

By construction, $D \subset \mathrm{Span}(B_k)$, so $V = \mathrm{Span}(D) \subset \mathrm{Span}(B) \subset V$. Thus $\mathrm{Span}(B_k) = V$. To prove that B_k is independent, we show more generally that each B_i is linearly independent. To see this, we proceed inductively. Note that B_0 is linearly independent. Assume that B_{i-1} is linearly independent. Then it is pretty straightforward to check that B_i is a minimal spanning set, hence linearly independent (check this!). Thus B_i is linearly independent, as needed. We conclude that B_k is a basis for V. $\square$

Exercise 1. Using Proposition 1.4, show that if $\dim(V) < \infty$ and $W \subseteq V$ then

$$\dim(W) \leq \dim(V).$$

In particular, if V is finite-dimensional, then any subspace of V is finite-dimensional.

Answer. Let B be a basis for W. Then by Proposition 1.4, there exists a set $B' \subset V$ such that $B \cup B'$ is a basis for V. Thus we have

$$\dim(W) = |B| \leq |B| + |B'| = \dim(V).$$

$\square$

Exercise 2. Using Proposition 1.4, show that if W is a subspace of V and $\dim(V) = \dim(W) < \infty$ then $V = W$.

Answer. Let B be a basis for W. Then by Proposition 1.4, there exists a set $B' \subset V$ such that $B \cup B'$ is a basis for V. Therefore,

$$|B| + |B'| = \dim(V) = \dim(W) = |B|.$$

It follows that $|B'| = 0$, i.e., $B = \emptyset$. Thus B is a basis for V. Since it is also a basis for W, we have $V = W$. $\square$

2 The v.s.-theoretic Definition of Rank; Nullity

Meyer defines the rank of a matrix A to be the number of pivots in any row echelon form of A (or equivalently, the number of pivots in E_A, the reduced echelon form of A).

Let us now introduce the vector space-theoretic definition of rank, as well as a similar definition of nullity. Once you get used to it, the v.s.-theoretic definition of rank is very geometrically intuitive and elegant. When we actually do computations of rank, it is useful to also have Meyer's definition in mind. At the end of Section 6, you will verify in the exercises that the two definitions are the same.

For any linear transformation $T : V \to W$ between vector spaces V and W, define

$$\operatorname{rank}(V) = \dim \operatorname{im}(V),$$

and define

$$\text{nullity}(V) = \dim\ker(V).$$

For A a matrix, define

$$\text{rank}(A) = \text{rank}(T_A),$$

and define

$$\text{nullity}(A) = \text{nullity}(T_A).$$

Note that by definition $\text{rank}(A) = \dim \mathcal{R}(A)$, the dimension of the range (i.e., column space) of A, and $\text{nullity}(A)$ is just the dimension of the space of solutions to the system $A\vec{x} = \vec{0}$.

Here is one of the central theorems of linear algebra:

Theorem 2.1 (Rank-Nullity Theorem)**.** *For $T : V \to W$ a linear transformation between vector spaces V and W with* $\dim(V) \leq \infty$*,*

$$\dim(V) = \text{rank}(T) + \text{nullity}(T).$$

Proof. By Exercise 1, we know $\ker(T)$ has a finite basis. Let $\{\mathbf{u}_1, \ldots, \mathbf{u}_m\}$ be a basis for $\ker(T)$. By Proposition 1.4, we can extend this to a basis for V:

$$\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_n\}.$$

Since $\text{nullity}(T) = \dim\ker(T) = m$ and $\dim(V) = m + n$, it suffices to show that $\text{rank}(T) = \dim \text{im}(T) = n$. To do this, we show that

$$\{T\mathbf{w}_1, \ldots, T\mathbf{w}_n\}$$

is a basis for $\text{im}(T)$.

Let v be an arbitrary vector in V. Then there exist scalars

$$a_1, \ldots, a_m, b_1, \ldots, b_n$$

such that:

$$\mathbf{v} = a_1\mathbf{u}_1 + \cdots + a_m\mathbf{u}_m + b_1\mathbf{w}_1 + \cdots + b_n\mathbf{w}_n.$$

Therefore,

$$\begin{aligned} T\mathbf{v} &= a_1T\mathbf{u}_1 + \cdots + a_mT\mathbf{u}_m + b_1T\mathbf{w}_1 + \cdots + b_nT\mathbf{w}_n \\ &= b_1T\mathbf{w}_1 + \cdots + b_nT\mathbf{w}_n, \end{aligned}$$

since each $T\mathbf{u}_i = 0$. Since every vector in $\operatorname{im} T$ is of the form Tv for some $v \in V$, this shows that $\operatorname{Span}\{T\mathbf{w}_1, \ldots, T\mathbf{w}_n\} = T$.

Now we check that $\{T\mathbf{w}_1, \ldots, T\mathbf{w}_n\}$ is linearly independent: Suppose we have scalars $c_1, \ldots, c_n$ with

$$c_1 T\mathbf{w}_1 + \cdots + c_n T\mathbf{w}_n = 0.$$

Then by linearity,

$$T(c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n) = 0.$$

That is,

$$c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n \in \ker(T).$$

Then, since the $\mathbf{u}_i$ form a basis for $\ker T$, there exist scalars d_i such that:

$$c_1\mathbf{w}_1 + \cdots + c_n\mathbf{w}_n = d_1\mathbf{u}_1 + \cdots + d_m\mathbf{u}_m.$$

Since $\{\mathbf{u}_1, \ldots, \mathbf{u}_m, \mathbf{w}_1, \ldots, \mathbf{w}_n\}$ is a basis for V, all the scalars c_i and d_i must be zero. Therefore, $\{T\mathbf{w}_1, \ldots, T\mathbf{w}_n\}$ is linearly independent and is thus a basis of for $\operatorname{im}(T)$, as desired. □

Exercise 3. Using Theorem 2.1, give a quick proof that the rank nullity theorem for matrices also holds: For A any matrix,

$$\#\text{ columns of A} = \operatorname{rank}(A) + \operatorname{nullity}(A).$$

[Note: Once we establish that the two definitions of rank are the same, the result of this exercise can also be deduced by observing that $\operatorname{rank}(A)$ is the number of basic columns of A, and $\operatorname{nullity}(A)$ is the number of non-basic columns of A; since every column is either basic or non-basic, the result follows. This is Meyer's argument in Section 4.4.]

Answer. By definition, $\operatorname{rank}(A) = \operatorname{rank}(T_A)$ and $\operatorname{nullity}(A) = \operatorname{nullity}(T_A)$. Let $n = \#$ columns of A. Then T_A is a linear map with domain $\mathbb{R}^n$. By the rank nullity theorem for linear maps, we have

$$\operatorname{rank}(A) + \operatorname{nullity}(A) = \operatorname{rank}(T_A) + \operatorname{nullity}(T_A) = n = \#\text{ columns of A},$$

as desired. □

3 Verifying that a Set of Vectors is a Basis in a Smarter Way

Proposition 3.1. *Let V be a vector space of dimension $n \leq \infty$ and let $S = \{v_1, v_2, \ldots, v_n\} \subset V$. The following are equivalent:*

1. *S is a basis for V,*

2. *S is linearly independent*

3. $\mathrm{Span}(S) = V$.

Proof. It's clear that from the definition of a basis that $1 \implies 3$ and $1 \implies 2$. Let's check that $2 \implies 1$ and $3 \implies 1$. The proof I showed in class uses the rank nullity theorem. A student in class suggested a different proof apprch, which I think is shorter and easier to follow than my proof from class. In this revision of the note, I'll share his approach.

First we show that $2 \implies 1$. If S is linearly independent, then it is a basis for $\mathrm{Span}(S)$. This implies that $\dim \mathrm{Span}(S) = n$. By Exercise 2, we must have $\mathrm{Span}(S) = V$, so S is a basis for V.

To show that $3 \implies 1$, suppose that $\mathrm{Span}(S) = V$. Then S contains a minimal spanning set S' for V. We've seen that a minimal spanning set is a basis, so S' is a basis for V. Since $\dim(V) = n$, it must be that $|S'| = n$, so in fact $S' = S$, which implies that S is a basis for V. □

Exercise 4. Taking advantage of Proposition 3.1, show that the columns of the matrix $A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$ form a basis for $\mathbb{R}^3$.

Answer. The row echelon form of A is $\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ 0 & 0 & -1 \end{pmatrix}$. This is a matrix with no non-basic columns, so the nullspace of A is $\{\vec{0}\}$. Thus, the columns of A are linearly independent. By Proposition 3.1 above, the columns of A form a basis for $\mathbb{R}^3$. □

I give another application of the rank-nullity theorem in Section 8 below.

4 Injections, Surections, Bijections, and Inverses

Let's review injective, surjective, and bijective functions. For any sets A, B, a function $f : A \to B$ is an **injection** (or is **1-1**) if whenever $f(a) = f(a')$, we have $a = a'$. f is a **surjection** (or **onto**) if the image of f is all of B; that is, for every $b \in B$ we have $f(a) = b$ for some $a \in A$. f is a **bijection** if f is both an injection and a surjection.

A function $g : B \to A$ is an **inverse** of f if

$$g \circ f = \mathrm{Id}_A,$$

$$f \circ g = \mathrm{Id}_B\,.$$

Exercise 5. Show that f has an inverse if and only if f is a bijection.

Answer. Suppose $f : A \to B$ has an inverse $g : B \to A$, so that we have

$$g \circ f = \mathrm{Id}_A,$$

$$f \circ g = \mathrm{Id}_B\,.$$

Suppose $f(a) = f(b)$. Then

$$a = g \circ f(a) = g \circ f(b) = b,$$

so f is injective. For $b \in B$, $b = f \circ g(b)$, i.e. $b = f(g(b))$ so f is surjective. Hence f is a bijection.

Conversely, suppose f is a bijection. By surjectivity, for each $b \in B$ there is some $b^* \in A$ with $f(b^*) = b$; by injectivity, that element b^* is unique. We may thus define $g : B \to A$ by taking $g(b) = b^*$ for all $b \in B$. It is easy to check that in fact $g \circ f = \mathrm{Id}_A$ and $f \circ g = \mathrm{Id}_B$, so g is an inverse for f. □

Exercise 6. Show that if f is a bijection, then its inverse is unique. [Hint: The argument is very similar the argument that the inverse of an invertible matrix is unique.]

Answer. Let $g, h : B \to A$ be two inverses for $f : A \to B$. We have

$$g = g \circ \mathrm{Id}_B = g \circ f \circ h = \mathrm{Id}_A \circ h = h.$$

□

We denote the inverse of f as f^{-1}.

Exercise 7. Show that if $f : A \to B$ is a bijection and $g : B \to C$ is a bijection, then $g \circ f : A \to C$ is a bijection.

Answer. It's easy to check that $g \circ f$ has inverse $f^{-1} \circ g^{-1}$. Since $g \circ f$ has an inverse, it is a bijection by Exercise 5. $\square$

5 Isomorphisms of Vector Spaces

We call a linear bijection of vector spaces an **isomorphism**. If there exists an isomorphism $T : V \to W$ between vector spaces V and W, we say V and W are **isomorphic.**

The word isomorphism derives from the Greek for "equal shape." Intuitively, if two vector spaces are isomorphic, they have the same "structure" as vector spaces. Even though they may be different as sets, isomorphic vector spaces are the same in all ways that matter in linear algebra.

Remark 5.1. The notion of isomorphism of objects is pervasive in mathematics and extends well beyond vector space—mathematicians talk about isomorphisms of all kinds of algebraic and geometric objects. The general setting for talking about isomorphisms of mathematical objects is called *category theory.*

Exercise 8. Show that if V and W are vector spaces and $T : V \to W$ is an isomorphism, then T^{-1} is also linear.

Answer. We need to show that for $w_1, w_2 \in W$ and $\alpha \in F$,

1. $T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2)$,

2. $T^{-1}(\alpha w_1) = \alpha T^{-1}(w_1)$.

Using the linearity of T, we have

$$T \circ T^{-1}(w_1 + w_2) = w_1 + w_2 = T \circ T^{-1}(w_1) + T \circ T^{-1}(w_2) = T(T^{-1}(w_1) + T^{-1}(w_2)).$$

T is an isomorphism, so is injective. Thus we have that

$$T^{-1}(w_1 + w_2) = T^{-1}(w_1) + T^{-1}(w_2).$$

The argument that $T^{-1}(\alpha w_1) = \alpha T^{-1}(w_1)$ is similar. $\square$

The following appeared as a bonus question on Quiz 6:

Proposition 5.2. *For $T : V \to W$ a linear map between vector spaces, prove that* $\ker(T) = \{\vec{0}\}$ *if and only if T is injective.*

Proof. If T is injective, then for every $w \in W$, there is at most one $v \in V$ with $T(v) = w$. Taking $w = \vec{0}$, we have that $\ker(T)$ contains at most one element. Thus to see that $\ker(T) = \{\vec{0}\}$, it suffices to observe that for any linear map T, $T(\vec{0}) = \vec{0}$: $T(\vec{0}) = T(\vec{0} + \vec{0}) = T(\vec{0}) + T(\vec{0})$, so by the cancellation law, $T(\vec{0}) = \vec{0}$. Thus $\ker(T) = \{\vec{0}\}$.

Conversely, suppose $\ker(T) = \{\vec{0}\}$. To establish injectivity of T, we need to show that for any v_1, v_2 with $T(v_1) = T(v_2)$, $v_1 = v_2$. Then

$$\vec{0} = T(v_1) - T(v_2) = T(v_1 - v_2),$$

so $v_1 - v_2 \in \ker(T)$. But since $\ker(T) = \{\vec{0}\}$, we have $v_1 - v_2 = \vec{0}$. Thus $v_1 = v_2$. □

Exercise 9. Show that if $T : V \to W$ is an isomorphism of finite-dimensional vector spaces and $S \subset V$, then

(a) S is linearly independent if and only if $T(S)$ is linearly independent.

(b) S is a basis for V if and only if $T(S)$ is a basis for W.

(c) Conclude that if V and W are isomorphic vector spaces of finite dimension then $\dim(V) = \dim(W)$.

[Note: The finite-dimension assumption is actually unnecessary here. That is to just make your life a little simpler.]

Answer. Let $S = \{s_1, s_k, \ldots, s_k\}$.

To prove (a), Suppose S is linearly independent, and let us consider a linear combination

$$\vec{0} = c_1 T(s_1) + \cdots + c_k T(s_k).$$

We need to show that $c_i = 0$ for each i. By linearity, we have

$$\vec{0} = T(c_1(s_1) + \cdots + c_k(s_k)).$$

Since T is an isomorphism, $\ker(T) = \{\vec{0}\}$, so

$$c_1 s_1 + \cdots + c_k s_k = \vec{0}.$$

Since S is linearly independent, we have that $c_i = 0$ for each i, as we wanted to show.

Conversely, suppose $T(S)$ is linearly independent. Then since T^{-1} is an isomorphism, the argument above gives that $T^{-1} \circ T(S) = S$ is linearly independent as well.

given (a), to prove (b), it suffices to show that $\mathrm{Span}(S) = V$ if and only if $\mathrm{Span}(T(S)) = W$. Because T is surjective, for any $w \in W$, there exists $v \in V$ with $T(v) = w$. If $\mathrm{Span}(S) = V$ then we may write

$$v = c_1 s_1 + \cdots + c_k s_k.$$

Then by the linearity of T,

$$w = c_1 T(s_1) + \cdots + c_k T(s_k).$$

Thus $w \in \mathrm{Span}(T(S))$. Since w was an arbitrary vector in W, this shows that $\mathrm{Span}(T(S)) = W$.

Conversely, if $\mathrm{Span}(T(S)) = W$, then since T^{-1} is an isomorphism, the argument above gives that $\mathrm{Span}(T^{-1} \circ T(S)) = \mathrm{Span}(S) = V$.

(c) $|T(S)| = |S|$, since T is a bijection. So from (b) we have

$$\dim(V) = |S| = |T(S)| = \dim(W).$$

□

Exercise 10. As a converse to the previous exercise, show that if V and W are vector spaces with $\dim(V) = \dim(W) \leq \infty$, then V and W are isomorphic. [Hint: consider an arbitrary bijection between any bases for V and W and extend this to an isomorphism.]

Answer. Let $B = \{b_1, \ldots, b_k\}$ and $D = \{d_1, \ldots, d_k\}$ be bases for V and W respectively. We've seen that any v can be written as

$$v = v_1 b_1 + \cdots + v_k b_k$$

for unique scalars $v_1, \ldots, v_k$. Thus we obtain a well-defined function $T : V \to W$ by taking

$$T(v_1 b_1 + \cdots + v_k b_k) = v_1 d_1 + \cdots + v_k d_k.$$

T is invertible; $T^{-1} : W \to V$ is given by

$$T{-}1(w_1 d_1 + \cdots + w_k d_k) = (w_1 b_1 + \cdots + w_k b_k).$$

It remains to check that T is linear. For $v, v' \in V$, write $v = \sum_{i=1}^{k} v_i b_i$ and $v' = \sum_{i=1}^{k} v'_i b_i$.

$$T(v) + T(v') = \sum_{i=1}^{k} v_i d_i + \sum_{i=1}^{k} v'_i d_i = \sum_{i=1}^{k} (v_i + v'_i) d_i = T(v + v').$$

The check that $T(cv) = cT(v)$ for all scalars c is essentially the same. □

Exercise 11.

(a) Show that for any $n \geq 0$, there exists an isomorphism between $\mathbb{R}^{n+1}$ and the vector space P of polynomials of degree at most n.

(b) Show that for any $m, n \geq 1$, there exists an isomorphism between $\mathbb{R}^{mn}$ and $M^{m \times n}$, the vector space of $m \times n$ matrices with real coefficients.

answer. In view of Exercise 10, to prove (a) it suffices to find a basis of size $n + 1$ for P. In fact,

$$S = \{1, x, x^2, \dots, x^n\}$$

is such a basis. Clearly, $\mathrm{Span}(S) = P$. S is also linearly independent, by the fundamental theorem of algebra, which tells us that a non-zero polynomial of degree n has at most n -distinct roots.

To prove (b), let $S^{m \times n}$ denote the set of $m \times n$ matrices with exactly one non-zero entry, equal to 1. $S^{m \times n}$ is clearly a basis for $M^{m \times n}$. Since $|S^{m \times n}| = mn = |S^{n \times m}|$, the result follows from Exercise 10. □

6 Reconciling the v.s.-theoretic Definition of Rank with Meyer's Definition

Proposition 6.1. *For linear maps $F : U \to V$, $G : V \to W$ with U, V, W, finite dimensional,*

1. $\mathrm{rank}(G \circ F) \leq \mathrm{rank}(G)$

2. $\mathrm{rank}(G \circ F) \leq \mathrm{rank}(F)$.

Proof. By the definition of rank we have that

$$\begin{aligned}\operatorname{rank}(G \circ F) &= \dim \operatorname{im}(G \circ F)\\ \operatorname{rank}(F) &= \dim \operatorname{im}(F)\\ \operatorname{rank}(G) &= \dim \operatorname{im}(G).\end{aligned}$$

To show 1, it suffices to show that $\operatorname{im}(G \circ F) \subseteq \operatorname{im}(G)$, by Exercise 1. But this is clear, because

$$\begin{aligned}\operatorname{im}(G) &= \{G(v) \mid v \in V\}\\ \operatorname{im}(G \circ F) &= \{G(v) \mid v \in \operatorname{im}(F)\},\end{aligned}$$

and $\operatorname{im}(F) \subseteq V$.

To show 2, note that the rank-nullity theorem tells us that

$$\begin{aligned}\operatorname{rank}(G \circ F) &= \dim(U) - \operatorname{nullity}(G \circ F),\\ \operatorname{rank}(F) &= \dim(U) - \operatorname{nullity}(F).\end{aligned}$$

Thus, it suffices to show that $\operatorname{nullity}(G \circ F) \geq \operatorname{nullity}(F)$. The rest of the argument is similar to the argument for 1. By definition,

$$\begin{aligned}\operatorname{nullity}(G \circ F) &= \dim \ker(G \circ F)\\ \operatorname{nullity}(F) &= \dim \ker(F),\end{aligned}$$

so by Exercise 1, it's enough to show that $\ker(F) \subseteq \ker(G \circ F)$. But this is clearly true because for any $u \in U$ with $F(u) = \vec{0}$, we also have $G \circ F(u) = \vec{0}$. This establishes 2, and thus completes the proof. □

Proposition 6.2. *If $E : V \to V$ and $G : W \to W$ are isomorphisms of finite-dimensional vector spaces and $F : V \to W$ is a linear map, then*

$$\operatorname{rank}(G \circ F \circ E) = \operatorname{rank}(F).$$

Proof. It suffices to show that

$$\operatorname{rank}(G \circ F \circ E) \leq \operatorname{rank}(F)$$

and

$$\operatorname{rank}(F) \leq \operatorname{rank}(G \circ F \circ E).$$

By Proposition 6.1, we have

$$\operatorname{rank}((G \circ F) \circ E) \leq \operatorname{rank}(G \circ F) \leq \operatorname{rank}(F).$$

Note that $F = G^{-1} \circ G \circ F \circ E \circ E^{-1}$. Thus we have

$$\operatorname{rank}(F) = \operatorname{rank}((G^{-1} \circ G \circ F \circ E) \circ E^{-1}) \leq \operatorname{rank}(G^{-1} \circ (G \circ F \circ E)) \leq \operatorname{rank}(G \circ F \circ E).$$

This gives the result. $\square$

Exercise 12. Prove that a square matrix A is invertible if and only if T_A is an isomorphism. [Hint: Recall that $T_{I_n} = \mathrm{Id}_{\mathbb{R}^n}$, the identity map on $\mathbb{R}^n$.]

Answer. If A is an invertible $n \times n$ matrix, we have

$$AA^{-1} = I_n,$$
$$A^{-1}A = I_n,$$

where I_n denotes the $n \times n$ identity matrix. Applying T to these equations, and using the fact that $T_C \circ T_D = T_{CD}$ for all matrices A, A^{-1} of the appropriate dimensions, we have

$$T_A \circ T_{A^{-1}} = T_{AA^{-1}} = \mathrm{Id}_{\mathbb{R}^n}$$
$$T_{A^{-1}} \circ T_A = T_{A^{-1}A} = \mathrm{Id}_{\mathbb{R}^n},$$

so T_A is an isomorphism with inverse $T_{A^{-1}}$.

Conversely, if T_A is an isomorphism, then we have

$$T_A \circ T_A^{-1} = \mathrm{Id}_{\mathbb{R}^n},$$
$$T_A^{-1} \circ T_A = \mathrm{Id}_{\mathbb{R}^n}.$$

Applying the "bracket map" to these equations, and using properties of T and $[\cdot]$ established in class a while ago, we have

$$A[T_A^{-1}] = [T_A][T_A^{-1}] = [T_A \circ T_A^{-1}] = I_n,$$
$$[T_A^{-1}]A = [T_A^{-1}][T_A] = [T_A^{-1} \circ T_A] = I_n.$$

so A is invertible, with $A^{-1} = [T_A^{-1}]$. $\square$

Exercise 13. Using Exercise 12 and Propsition 6.2, show that if A, B, and C are matrices of dimensions $m \times m$, $m \times n$, and $n \times n$, respectively, with A and C non-singular, then

$$\operatorname{rank}(ABC) = \operatorname{rank}(B).$$

Answer.

$$\operatorname{rank}(ABC) = \operatorname{rank}(T_{ABC}) = \operatorname{rank}(T_A \circ T_B \circ T_C) = \operatorname{rank}(T_B) = \operatorname{rank}(B);$$

the first and last equals follows from the definition of the rank of a matrix, and the third inequality follows from Exercise 12 and Propsition 6.2. $\square$

Exercise 14. Using Exercise 13, explain why the rank of a matrix A, as defined using the v.s.-theoretic definition, doesn't change when we perform elementary row and column operations on A.

Answer. Any sequence of elementary row and column operations can be performed on A by multiplying A on the left and right, respectively, by products of elementary matrices. Elementary matrices are non-singular, and so their products are non-singular as well. Thus the result follows from Exercise 13. $\square$

Exercise 15. Show that for a matrix A in reduced echelon form, Meyer's definition of rank(A) and the v.s.-theoretic definition of rank(A) coincide. Using exercise 14, conclude that for arbitrary matrices A, the two definitions coincide. (For this, we only need to think about row operations, in fact.)

Answer.

$$\begin{aligned}
\operatorname{rank}(E_A) &= \dim(\mathcal{R}(E_A)) \\
&= \dim \operatorname{Span}(\text{columns of } E_A) \\
&= \dim \operatorname{Span}(\text{basic columns of } E_A) \\
&= \#\text{ of basic columns of } E_A,
\end{aligned}$$

where the third equality holds because the non-basic columns of E_A are linear combinations of the basic columns of E_A; and last equality holds because the basic columns of E_A are standard basis vectors, so are linearly independent.

Since E_A is obtained from A (via Gauss-Jordan elimination) by performing elementary row operations on A, we have $\operatorname{rank}(A) = \operatorname{rank}(E_A)$ by Exercise 14. Thus $\operatorname{rank}(A) = \#$ of basic columns of E_A. This shows that our vector space definition of rank and Meyer's definition coincide. $\square$

7 Proving that $\operatorname{rank}(A) = \operatorname{rank}(A^T)$

One of the very fundamental facts from basic linear algebra is that for any matrix A, $\operatorname{rank}(A) = \operatorname{rank}(A^T)$. The proof is not quite as obvious as you might expect, but with the machinery we have now, it's reasonably easy. Meyer also gave a proof of this in 3.9, but I think it is more pleasant to treat this this question with the v.s.-theoretic definition of rank in mind, so I've waited on this.

First a lemma, which Meyer also established in Chapter 3.

Lemma 7.1. *any matrix $m \times n$ has a factorization of the form $A = U\Sigma V$, where U and V are non-singular matrices of dimension m and n, and Σ is a block matrix of the form*

$$\Sigma = \begin{pmatrix} I_r & \mathbf{0}^{r\times(n-r)} \\ \mathbf{0}^{(m-r)\times r} & \mathbf{0}^{(m-r)\times(n-r)} \end{pmatrix}$$

for some r; Here $\mathbf{0}^{a\times b}$ denotes the $a \times b$ matrix consisting of all zeros.

Proof. To put A in reduced echelon form E_A, we perform row operations. By doing column operation of type III on E_A, we can zero out the non-basic columns of E_A; then, via column operations of type I, we can move the basic columns of the resulting matrix into the first r positions, giving a matrix Σ as above. Thus we can transform A into Σ by performing row and column operations on A. The row operations can be carried out by multiplying A on the left by a non-singular matrix U'. The column operations can be carried out by multiplying A on the right by a non-singular matrix $V' : \Sigma = U'AV'$. Taking $U = U'^{-1}$, $V = V'^{-1}$, we then have $A = U\Sigma V$. □

Theorem 7.2. *For A any matrix, $\operatorname{rank}(A) = \operatorname{rank}(A^T)$.*

Proof. Consider the decomposition of $A = U\Sigma V$ as in Lemma 7.1. Clearly $\operatorname{rank}(\Sigma) = r$, so by Exercise 13, $\operatorname{rank}(A) = r$. Now consider the transpose: $A^T = V^T\Sigma^T V^T$.

$$\Sigma^T = \begin{pmatrix} I_r & \mathbf{0}^{r\times(m-r)} \\ \mathbf{0}^{(n-r)\times r} & \mathbf{0}^{(n-r)\times(m-r)} \end{pmatrix},$$

so clearly we also have that $\operatorname{rank}(\Sigma^T) = r$. The transpose of a nonsingular matrix is non-singular (do you remember why?), so by Exercise 13 again, we have that

$$\operatorname{rank}(A^T) = r = \operatorname{rank}(A).$$

□

8 A Loose End on the Inverses of Matrices

In 3.7, Meyer proved the following

Proposition 8.1. *If A and B are square matrices of the same dimension n and $AB = I_n$ then A and B are inverses; that is, $BA = I_n$, as well.*

I think this is a very cool and fundamental fact, and it is needed to show that we can compute the inverse on an $n \times n$ matrix by solving n systems of linear equations. But I found Meyer's proof ugly, so I skipped it in class. Now, we are ready for a slick proof.

Proof. First we check that A is invertible.

$$T_A \circ T_B = T_{AB} = T_{I_n},$$

so T_A must be surjective (check this). Thus

$$\operatorname{rank}(T_A) = \dim \operatorname{im}(T_A) = \dim \mathbb{R}^n = n.$$

By the rank-nullity theorem, then, $\operatorname{nullity}(T_A) = 0$, i.e. $\ker(T_A) = \{0\}$. By Proposition 5.2 then, T_A is injective. Thus T_A an isomorphism. By Exercise 11, then, A is invertible.

We've seen that the inverse of a matrix is unique, if it exists, so A has a unique inverse A^{-1}. Multiplying the equation $AB = I_n$ on the left by A^{-1} gives that $B = A^{-1}$, as desired. $\square$