Linearity

Paul Schrimpf
October 24, 2018
University of British Columbia
Economics 526

These notes are about linear algebra. References from the primary texts are chapters 10, 11, and 27 of Simon and Blume (1994), chapter 3 of De la Fuente (2000), and section 1.4 and portions of chapter 3 of Carter (2001). There are many mathematics texts on linear algebra. Axler (1997) is good. Many people like Gilbert Strang's video lectures (and his textbook). These notes were originally based on the material in Simon and Blume (1994), but they are now closer to the approach of Axler (1997) for finite dimensional spaces, and some mix of Luenberger (1969) and Clarke (2013) for infinite dimensional spaces.

We will study linear algebra with two goals in mind. First, we will finally carefully prove that the Lagrangian works. Recall that for a constrained optimization problem,

$$\max_x f(x) \text{ s.t. } h(x) = c,$$

we argued that x^* is a local max if for all v,

$$Dh_{x^*}v = 0 \implies Df_{x^*}v = 0.$$

We then made a heuristic argument that this is equivalent to the existence of Lagrange multipliers such that

$$Df_{x^*} + \lambda^T Dh_{x^*} = 0.$$

This result will be a consequence of the separating hyperplane theorem (or the geometric form of the Hahn-Banach theorem for infinite dimensional spaces).

The second main result that we will build toward are the first and second welfare theorems. The first welfare theorem states that a competitive equilibrium is Pareto efficient. The second welfare theorems states that every Pareto efficient allocation can be achieved by some competitive equilibrium. The second welfare theorem is also a consequence of the separating hyperplane theorem. The welfare theorems involve preferences (the subject of the notes on sets), vector spaces (the topic of these notes), and some continuity (the subject of the previous set of notes).

1. Vector spaces

A vector space is a set whose elements can be added and scaled. Vector spaces appear quite often in economics because many economic quantities can be added and scaled. For example, if firm A produces quantities y_1^A and y_2^A of goods 1 and 2, while firm B

produces (y_1^B, y_2^B), then total production is $(y_1^A + y_1^B, y_2^A + y_2^B)$. If firm A becomes 10% more productive, then it will produce $(1.1y_1^A, 1.1y_2^A)$.

We have been working with $\mathbb{R}^n$, which is the most common vector space. There are three ways of approaching vector spaces. The first is geometrically — introduce vectors as directed arrows. This works well in $\mathbb{R}^2$ and $\mathbb{R}^3$ but is difficult in higher dimensions. The second is analytically — by treating vectors as n-tuples of numbers $(x_1, ..., x_n)$. The third approach is axiomatically — vectors are elements of a set that has some special properties. You likely already have some familiarity with the first two approaches. Here, we are going to take the third approach. This approach is more abstract, but this abstraction will allow us to generalize what we might know about $\mathbb{R}^n$ to other more exotic vector spaces. Also, some theorems and proofs become shorter and more elegant.

Definition 1.1. A **vector space** is a set V with two operations, addition $+$, which takes two elements of V and produces another element in V, and scalar multiplication $\cdot$, which takes an element in V and an element in $\mathbb{R}$ and produces an element in V, such that

(1) $(V, +)$ is a commutative group, i.e.
   (a) Closure: $\forall v_1 \in V$ and $v_2 \in V$ we have $v_1 + v_2 \in V$.
   (b) Associativity: $\forall v_1, v_2, v_3 \in V$ we have $v_1 + (v_2 + v_3) = (v_1 + v_2) + v_3$.
   (c) Identity exists: $\exists 0 \in V$ such that $\forall v \in V$, we have $v + 0 = v$
   (d) Invertibility: $\forall v \in V \; \exists -v \in V$ such that $v + (-v) = 0$
   (e) Commutativity: $\forall v_1, v_2 \in V$ we have $v_1 + v_2 = v_2 + v_1$
(2) Scalar multiplication has the following properties:
   (a) Closure: $\forall v \in V$ and $\alpha \in \mathbb{R}$ we have $\alpha v \in V$
   (b) Distributivity: $\forall v_1, v_2 \in V$ and $\alpha_1, \alpha_2 \in \mathbb{R}$

$$\alpha_1(v_1 + v_2) = \alpha_1 v_1 + \alpha_1 v_2$$

   and

$$(\alpha_1 + \alpha_2)v_1 = \alpha_1 v_1 + \alpha_2 v_1$$

   (c) Consistent with field multiplication: $\forall v \in V$ and $\alpha_1, \alpha_2 \in V$ we have

$$1v = v$$

   and

$$(\alpha_1 \alpha_2)v = \alpha_1(\alpha_2 v)$$

A vector space is also called a linear space. Like Carter (2001) says, "This long list of requirements does not mean that a linear space is complicated. On the contrary, linear spaces are beautifully simple and possess one of the most complete and satisfying theories in mathematics. Linear spaces are also immensely useful providing one of the principal foundations of mathematical economics. The most important examples of linear spaces are $\mathbb{R}$ and $\mathbb{R}^n$. Indeed, the abstract notion of linear space generalizes the algebraic behavior of $\mathbb{R}$ and $\mathbb{R}^n$." One way of looking at vector spaces is that they are a way of trying to generalize the things that we know about two and three dimensional space to other contexts.

1.0.1. *Examples.* We now give some examples of vector spaces.

Example 1.1. $\mathbb{R}^n$ is a vector space. You are likely already familiar with this space. Vector addition and multiplication are defined in the usual way. If $\mathbf{x}_1 = (x_{11}, ..., x_{n1})$ and $\mathbf{x}_2 = (x_{12}, ..., x_{n2})$, then vector addition is defined as

$$\mathbf{x}_1 + \mathbf{x}_2 = (x_{11} + x_{12}, ..., x_{n1} + x_{n2}).$$

The fact that $(\mathbb{R}^n, +)$ is a commutative group follows from the fact that $(\mathbb{R}, +)$ is a commutative group. Scalar multiplication is defined as

$$a\mathbf{x} = (ax_1, ..., ax_n)$$

for $a \in \mathbb{R}$ and $\mathbf{x} \in R^n$. You should verify that the three properties in the definition of vector space hold. The vector space $(\mathbb{R}^n, \mathbb{R}, +, \cdot)$ is so common that it is called **Euclidean space**[a] We will often just refer to this space as $\mathbb{R}^n$, and it will be clear from context that we mean the vector space $(\mathbb{R}^n, \mathbb{R}, +, \cdot)$. In fact, we will often just write V instead of $(V, \mathbb{R}, +, \cdot)$ when referring to a vector space.

[a] To be more accurate, Euclidean space refers to $\mathbb{R}^n$ as an inner product space, which is a special kind of vector space that will be defined below.

Example 1.2. The set of all solutions to a homogenous system of linear equation with the right hand size equal to 0, i.e., $(x_1, ..., x_n) \in \mathbb{R}^n$ such that

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + ... + a_{1n}x_n &= 0 \\ a_{21}x_1 + a_{22}x_2 + ... + a_{2n}x_n &= 0 \\ &\vdots \\ a_{m1}x_1 + a_{m2}x_2 + ... + a_{mn}x_n &= 0, \end{aligned}$$

Most of the time, the two operations on a vector space are the usual addition and multiplication. However, they can appear different, as the following example illustrates.

Example 1.3. Take $V = \mathbb{R}^+$. Define "addition" as $x \oplus y = xy$ and define "scalar multiplication" as $\alpha \odot x = x^\alpha$. Then $(\mathbb{R}^+, \mathbb{R}, \oplus, \odot)$ is a vector space with identity element 1.

The previous few examples are each finite dimensional vector spaces. There are also infinite dimensional vector spaces.

Example 1.4. Let $V = \{$ all sequences of real numbers $\}$. For two sequences $\mathbf{x} = \{x_1, x_2, ...\}$ and $\mathbf{y} = \{y_1, y_2, ...\}$, define $\mathbf{x} + \mathbf{y} = \{x_1 + y_1, x_2 + y_2, ...\}$ and defie scalar multiplication as $\alpha\mathbf{x} = \{\alpha x_1, \alpha x_2, ...\}$. Then this is a vector space.

We encounter vector spaces of sequences in economics when we study infinite horizon discrete time optimization problems.

Spaces of functions are often vector spaces. In economic theory, we might want to work with a set of functions because we want to prove something for all functions in the set.

That is, we prove something for all utility functions or for all production functions. In non-parametric econometrics, we try to estimate an unknown function instead of an unknown finite dimensional parameter. For example, instead of linear regression $y = x\beta + \epsilon$ where want to estimate the unknown vector β, we might say $y = f(x) + \epsilon$ and try to estimate the unknown function f.

Here are some examples of vector spaces of functions. It would be a good exercise to verify that these examples have all the properties listed in the definition of a vector space.

Example 1.5. Let V = all functions from $[0, 1]$ to $\mathbb{R}$. For $f, g \in V$, define $f + g$ by $(f + g)(x) = f(x) + g(x)$. Define scalar multiplication as $(\alpha f)(x) = \alpha f(x)$. Then this is a vector space.

Sets of functions with certain properties also form vector spaces.

Example 1.6. The set of all continuous functions with addition and scalar multiplication defined as in 1.5.

Example 1.7. The set of all k times continuously differentiable functions with addition and scalar multiplication defined as in 1.5.

Example 1.8. The set of all polynomials with addition and scalar multiplication defined as in 1.5.

Example 1.9. The set of all polynomials of degree at most d with addition and scalar multiplication defined as in 1.5.

Generally, the vector space with which we are most interested is Euclidean space, $\mathbb{R}^n$. In fact, a good way to think about other vector spaces is that they are just variations of $\mathbb{R}^n$. The whole reason for defining and studying abstract vector spaces is to take our intuitive understanding of two and three dimensional Euclidean space and apply it to other contexts. If you find the discussion of abstract vector spaces and their variations to be confusing, you can often ignore it and think of two or three dimensional Euclidean space instead.

Vector spaces often contain other vector spaces. For example, either axis (or more generally any line passing through the origin) in $\mathbb{R}^2$ is itself a vector space.

Definition 1.2. A set $S \subseteq V$ is called a **linear subspace** if it is closed under (i) scalar multiplication and (ii) addition in other words, if

(i) for every $\mathbf{x} \in S$ and $\alpha \in \mathbb{R}$, we have $\alpha\mathbf{x} \in S$, and
(ii) for every $\mathbf{x} \in S$ and $\mathbf{y} \in S$, we have $\mathbf{x} + \mathbf{y} \in S$

This two requirements are sometimes written more succintly as $\forall \mathbf{x}, \mathbf{y} \in S$ and $\alpha \in \mathbb{R}$, $\alpha\mathbf{x} + \mathbf{y} \in S$. When the "linear" is clear from context, linear subspaces are often simply called subspaces.

Exercise 1.1. Show that if S is a linear subspace, then $0 \in S$.

Example 1.10. For any vector space V, the set of containing only the 0 element, $\{0\}$, is a linear subspace.

Example 1.11. In $\mathbb{R}^2$, any line passing through the origin is a linear subspace. In $\mathbb{R}^3$, any line passing through the origin, and any plane passing through the origin is a linear subspace.

Example 1.12. The set of all continuous functions from $\mathbb{R} \to \mathbb{R}$ such that $f(294) = 0$ is a linear subspace of the vector space of continuous functions.

Example 1.13. The set of all polynomials of degree at most d is a linear subspace of the space of all polynomials.

The intersection of two linear subspaces is also a linear subspace.

Example 1.14. If S and W are linear subspaces of V, then so is $S \cap W$.

However, the union of two linear subspaces is not necessarily a subspace. For example in $\mathbb{R}^2$, the two axes are each subspaces, but their union is not because, $(1, 0)$ and $(0, 1)$ are both in the union, but $(1, 0) + (0, 1) = (1, 1)$ is not.

Definition 1.3. Let $S_1, \ldots, S_k$ be linear subspaces of V. The **sum** of them is

$$S_1 + \cdots + S_k = \{\mathbf{x}_1 + \cdots + \mathbf{x}_k : \mathbf{x}_j \in S_j\}$$

$S_1 + \cdots + S_k$ is another linear subspace of V.

1.1. Linear combinations.

Definition 1.4. Let V be a vector space and $\mathbf{x}_1, \ldots, \mathbf{x}_k \in V$. A **linear combination** of $\mathbf{x}_1, \ldots, \mathbf{x}_k$ is any vector

$$c_1\mathbf{x}_1 + \ldots + c_k\mathbf{x}_k$$

where $c_1, \ldots, c_k \in \mathbb{R}$.

Note that by the definition of a vector space (in particular the requirement that vector spaces are closed under addition and multiplication), it must be that $c_1\mathbf{x}_1 + \ldots + c_k\mathbf{x}_k \in V$.

If we take all possible linear combinations of, $\{c_1x_1 + c_2x_2 : c_1 \in \mathbb{R}, c_2 \in \mathbb{R}\}$, then the set will contain 0, and it will be a linear subspace. This motivates the following definition.

Definition 1.5. Let V be a vector space and $W \subseteq V$. The **span** of W is the set of all finite linear combinations of elements of W

When W is finite, say $W = \{\mathbf{x}_1, \ldots, \mathbf{x}_k\}$, the span of W is the set

$$\{c_1\mathbf{x}_1 + \ldots + c_k\mathbf{x}_k : c_1, \ldots, c_k \in \mathbb{R}\}.$$

When W is infinite, the span of W is the set of all finite weighted sums of elements of W.

Lemma 1.1. *The **span** of any $W \subseteq V$ is a linear subspace.*

Proof. Left as an exercise. □

Example 1.15. Let V be the vector space of all functions from $[0,1]$ to $\mathbb{R}$ as in example 1.5. The span of $\{1, x, ..., x^n\}$ is the set of all polynomials of degree less than or equal n.

1.2. **Dimension, linear independence, and basis.** You are probably familiar with the idea that $\mathbb{R}^n$ is n-dimensional. Roughly speaking if a vector space is n dimensional, then we should be able to describe any vector in it by listing n scalars or coordinates. In this section we formally define dimension.

Definition 1.6. A set of vectors $W \subseteq V$, is **linearly independent** if the only solution to

$$\sum_{j=1}^{k} c_j \mathbf{x}_j = 0$$

is $c_1 = c_2 = ... = c_k = 0$ for any k and $\mathbf{x}_1, ..., \mathbf{x}_k \in W$. If W is not linearly independent, then it is **linearly dependent**,

Example 1.16. In $\mathbb{R}^2$, $\{(1,0),(0,1)\}$ is linearly independent. $\{(0,0)\}$ is linearly dependent. Any set of three or more vectors in linearly dependent.

Definition 1.7. The **dimension** of a vector space, V, is the cardinality of the largest set of linearly independent elements in V.

How large can the largest set of linearly independent elements be? The following theorem is a starting point. It tells us that a linearly independent set is smaller than any set that spans.

Comment 1.1. The above definitions for linear independence and dimension work for any dimension, finite or infinite. However, some of the theorems that follow are awkward to prove for infinite dimension, so some of them will assume finite dimension. Exercise 1.3, sketches how to prove the existence of a basis for infinite dimensional spaces.

Theorem 1.1. *Suppose $\mathbf{v}_1, ..., \mathbf{v}_n$ span V, and $\mathbf{u}_1, ..., \mathbf{u}_m$ are linearly independent, then $n \geq m$.*

Proof. Since $\mathbf{v}_1, ..., \mathbf{v}_n$ span V, it must be that $\mathbf{u}_1, \mathbf{v}_1, ..., \mathbf{v}_n$ are linearly dependent. Therefore, $\exists \alpha_j \in \mathbb{R}$ and not all 0 such that

$$\alpha_1 \mathbf{u}_1 + \sum_{j=1}^{n} \alpha_{j+1} \mathbf{v}_j = 0$$

Since the $\mathbf{u}$'s are linearly independent, $\mathbf{u}_1 \neq 0$. Therefore, for some $j \geq 2$, $\alpha_j \neq 0$. Let ℓ be the largest such j. We can then rearrange to write

$$\mathbf{v}_\ell = -\frac{\alpha_1}{\alpha_\ell}\mathbf{u}_1 - \sum_{j=1}^{\ell-1} \frac{\alpha_{j+1}}{\alpha_\ell}\mathbf{v}_j = 0.$$

It follows that we can remove $\mathbf{v}_\ell$ and the remaining $\mathbf{v}$'s along with $\mathbf{u}_1$ will still span V.

We can then repeat the above argument. Since $\mathbf{u}_1, ..., \mathbf{u}_{i-1}, \mathbf{v}_1, ..., \mathbf{v}_{n-i+1}$ span V, it must be that $\mathbf{u}_1, ..., \mathbf{u}_i, \mathbf{v}_1, ..., \mathbf{v}_{n-i+1}$ are linearly dependent. Therefore, $\exists \alpha_j \in \mathbb{R}$ and not all 0 such that

$$\sum_{k=1}^{i} \alpha_i \mathbf{u}_i + \sum_{j=1}^{n} \alpha_{j+1} \mathbf{v}_j = 0$$

Since the $\mathbf{u}$'s are linearly independent, for some $j \geq i+1$, $\alpha_j \neq 0$. Let ℓ be the largest such j. We can then rearrange to write

$$\mathbf{v}_\ell = -\sum_{k=1}^{i} \frac{\alpha_i}{\alpha_\ell} \mathbf{u}_i - \sum_{j=1}^{\ell-1} \frac{\alpha_{j+i+1}}{\alpha_\ell} \mathbf{v}_j = 0.$$

It follows that we can remove $\mathbf{v}_\ell$ and the remaining $\mathbf{v}$'s along with $\mathbf{u}_1, ..., \mathbf{u}_i$ will still span V.

If there are fewer $\mathbf{v}$'s than $\mathbf{u}$'s, then the above induction contradicts the assumption that the $\mathbf{u}$'s are linearly independent. □

Comment 1.2. A more general version of theorem 1.1 is that if $W \subseteq V$ spans V and $U \subseteq V$ is linearly independent, then $|W| \geq |U|$.

The argument we used to prove theorem 1.1 does not suffice when U is infinite because the iterative replacement of elements of W with elements of U would never terminate. Exercise 1.4 sketches how to prove this theorem with infinite U.

This theorem implies that if B is linearly independent and spans V, then any other linearly independent set must have smaller cardinality. Hence, the dimension of V must equal the cardinality of B. Sets that are linearly independent and span a vector space are very useful, so they have a name.

Definition 1.8. A **basis** of a vector space V is any set of linearly independent vectors B such that the span of B is V.

If V has a basis with k elements, then the dimension of V must be at least k. In fact, the previous theorem implies that dimension of V must be exactly k. Another consequence is that any two bases must have the same cardinality.

Corollary 1.1. *Any two bases for a vector space have the same cardinality.*

Proof. Let B_1 and B_2 be bases for a vector space V. Since B_1 spans and B_2 is linearly independent, by theorem 1.1, $|B_1| \geq |B_2|$. Conversely, since B_2 spans and B_1 is linearly independent, $|B_2| \geq |B_1|$. Hence $|B_1| = |B_2|$. □

Example 1.17. A basis for $\mathbb{R}^n$ is $e_1 = (1, 0, ..., 0)$, $e_2 = (0, 1, 0, ..., 0)$, ..., $e_n = (0, ..., 0, 1)$. This basis is called the standard basis of $\mathbb{R}^n$.

The standard basis is not the only basis for $\mathbb{R}^n$. In fact, there are infinite different bases. Can you give some examples?

Exercise 1.2. What is the dimension of each of the examples of vector spaces above? Can you find a basis for them?

Note that an important requirement for a basis is that every $\mathbf{x} \in V$ can be written as a **finite** sum of basis elements. Therefore, for example, in $\ell^\infty = \{(x_1, x_2, ...) : x_i \in \mathbb{R}, \sup_{1 \leq i \leq \infty} |x_i| < \infty\}$, consider the set $E = \{e_i\}_{i=1}^\infty$, where e_i is an element of all 0's, except for a 1 in the ith position. E is linearly independent, but E does not span ℓ^∞ because e.g. you cannot write $(1, 1, 1, 1, ...)$ as a finite sum of the elements of E.

Given a set that spans a vector space, it is always possible to remove elements until the set is also linearly independent, and hence a basis. Doing this will be useful in various proofs, so we state it is a lemma.

Lemma 1.2. *Suppose* $\mathbf{v}_1, ..., \mathbf{v}_n$ *span* V. *Then there is a subset of the* $\mathbf{v}$*'s that is a basis for* V.

Proof. We proceed by induction. If $\mathbf{v}_1 = 0$, then remove it. Otherwise, keep it. For $j = 2, ..., n$, if $\mathbf{v}_j \in \text{span}(\mathbf{v}_1, ..., \mathbf{v}_{j-1})$, then delete it, otherwise keep it. At every step, the remaining $\mathbf{v}$ still span V. Furthermore, each step ensures that the non-deleted $\mathbf{v}_1, ..., \mathbf{v}_j$ are linearly independent. □

Conversely, every linearly independent set can be expanded to a basis.

Lemma 1.3. *Suppose* V *is finite dimensional and* $\mathbf{v}_1, ..., \mathbf{v}_n$ *are linearly independent. Then* $\exists$ $\mathbf{v}_{n+1}, ..., \mathbf{v}_m$ *such that* $\mathbf{v}_1, ..., \mathbf{v}_m$ *is a basis for* V.

Proof. By assumption V is finite dimensional, so it has a basis, say $\mathbf{u}_1, ..., \mathbf{u}_m$. Following the same argument as in the proof of theorem 1.1, the $\mathbf{u}$'s can be replaced by the $\mathbf{v}$'s to get another basis consisting of all the $\mathbf{v}$'s and $m - n$ of the $\mathbf{u}$'s. □

This lemma implies that if V is finite dimensional, then a basis for V exists. Even if V is infinite dimensional, then a basis exists. The argument is outlined in the following exercise.

Exercise 1.3. This exercise sketches how to prove the existence of a basis for infinite dimensional vector spaces.

A partial order is a relation that is reflexive, transitive, and antisymmetric. A set with a partial order is called a partially ordered set. Not all elements of a partially ordered set are comparable (a partial order need not be complete). A **chain** is any subset of a partially order set where all elements are comparable to one another. If Y is a partially ordered set with partial order $\geq$, and $A \subseteq Y$, then an upper bound for A is a $y \in Y$ such that $y \geq z$ for all $z \in A$. $y \in Y$ is a maximal element if there does not exist any $z \in Y$ such that $z > y$.

Zorn's lemma[a] says that if every chain in a partially ordered set has an upper bound, then the partially ordered set has a maximal element.

(1) Let X be a set $\mathcal{P}(X)$ be a the power set of X. Show that $\subseteq$ is a partial order on $\mathcal{P}(X)$.

(2) Let V be a vector space (possibly of infinite dimension) and let $L \subseteq V$ be linearly independent. Define

$$\mathcal{P} = \{S \subseteq V : L \subseteq S \text{ and } S \text{ linearly independent}\}$$

Show that $(\mathcal{P}, \subseteq)$ is a partially order set (this is short).

(3) Let $\mathcal{C} \subseteq \mathcal{P}$ be a chain. Show that $\cup_{C \in \mathcal{C}} C$ is an upper bound for $\mathcal{C}$. [Hint: you need to show $\cup_{C \in \mathcal{C}} C \in \mathcal{P}$.]

(4) Argue that $\mathcal{P}$ has a maximal element, B. [Hint: use Zorn's lemma.]

(5) Show that $\text{span}(B) = V$. [Hint: if $x \notin \text{span}(B)$, then $\{x\} \cup B \in \mathcal{P}$.]

(6) Conclude B is a basis for V.

[a] Zorn's lemma is equivalent to the axiom of choice, which is a basic assumption of set theory. The axiom of choice says that given a collection of non-empty sets, $\{S_i\}_{i \in \mathcal{I}}$, we can choose an element from each set $\{x_i\}_{i \in \mathcal{I}}$.

Exercise 1.4. This exercise shows that any two bases for a vector space must have the same cardinality. Let V be a vector space and B and A be bases for V.

(1) Argue that for each $b \in B$, $\exists$ finite $A_b \subseteq A$ such that $b = \sum_{a \in A_b} x_a a$ for some $x_a \in \mathbb{R}$.

(2) Argue that $A = \cup_{b \in B} A_b$

(3) Show that for any sets U and any infinite set I, if $U = \cup_{i \in I} F_i$ with each F_i finite, then $|U| \leq |I|$ (i.e. there exists a one-to-one (but not neccessarily onto) function from U to I).

(4) Conclude $|A| \leq |B|$, and reversing the roles of A and B gives $|B| \leq |A|$, so $|A| = |B|$.

The elements of a vector space can always be written uniquely in terms of a basis.

Lemma 1.4. *Let B be a basis for a vector space V. Then $\forall \mathbf{x} \in V$ there exists a unique $x_1, \ldots, x_k \in \mathbb{R}$ and $b_1, \ldots, b_k \in B$ such that $\mathbf{x} = \sum_{i=1}^{k} x_i b_i$*

Proof. By the definition of a basis, B spans V, so such $(x_1, \ldots, x_k)$ must exist. Now suppose there exists another such $(x'_1, \ldots, x'_j)$ and associated b'_i. The $\{b_1, \ldots, b_k\}$ and $\{b'_1, \ldots, b'_j\}$ might not be the same collection of elements of B. Let $\{\tilde{b}_1, \ldots, \tilde{b}_n\} = \{b_1, \ldots, b_k\} \cup \{b'_1, \ldots, b'_j\}$. Define $\tilde{x}_i = x_j$ if $\tilde{b}_i = b_j$, else 0. Similarly define $\tilde{x}'_i$. With this new notation we have

$$v = \sum_{i=1}^{n} \tilde{x}_i \tilde{b}_i = \sum_{i=1} \tilde{x}'_i \tilde{b}_i$$

$$\sum_{i=1}^{n} (\tilde{x}_i - \tilde{x}'_i) \tilde{b}_i = 0$$

However, if B is a basis, its elements must be linearly independent so $\tilde{x}_i = \tilde{x}'_i$ for all i, so the original $x_1, \ldots, x_k$ must be unique. □

1.3. $\mathbb{R}^n$ **as the only finite dimensional vector spaces.** $\mathbb{R}^n$ is the only n-dimesion vector space in the sense that any other finite dimensional vector space can be viewed as a simple change of basis.

Suppose V is an n-dimension vector space. By the definition of dimension, there must be a set of n linearly independent elements that span V. These elements form a basis. Call them $b_1, ..., b_n$. For each $\mathbf{x} \in V$, there are unique $x_1, ..., x_n \in \mathbb{R}$ such that

$$\mathbf{x} = \sum_{i=1}^{n} x_i b_i.$$

Thus we can construct a function, say $\mathcal{I} : V \to \mathbb{R}^n$ defined by

$$\mathcal{I}(\mathbf{x}) = (x_1, ..., x_n).$$

By lemma 1.4, $\mathcal{I}$ must be one-to-one. $\mathcal{I}$ must also be onto since by definition of a vector space, for any $(x_1, ..., x_n) \in \mathbb{R}^n$, the linear combination, $\sum_{i=1}^{n} x_i b_i$ is in V. Moreover, $\mathcal{I}$ preserves addition in that for any $\mathbf{x}^1, \mathbf{x}^2 \in V$,

$$\begin{aligned}
\mathcal{I}(\mathbf{x}^1 + \mathbf{x}^2) =& (x_1^1 + x_1^2, ..., x_n^1 + x_n^2) \\
=& (x_1^1, ..., x_n^1) + (x_1^2 + ... + x_n^2) \\
=& \mathcal{I}(\mathbf{x}^1) + \mathcal{I}(\mathbf{x}^2).
\end{aligned}$$

Similarly, $\mathcal{I}$ preserves scalar multiplication in that for all $\mathbf{x} \in V$, $\alpha \in \mathbb{R}$

$$\mathcal{I}(\alpha\mathbf{x}) = \alpha\mathcal{I}(\mathbf{x}).$$

Thus, V and $\mathbb{R}^n$ are essentially the same in that there is a one-to-one and onto mapping between them that preserves all the properties that make them vector spaces.

Definition 1.9. Let V and W be vector spaces over the field $\mathbb{F}$. V and W are **isomorphic** if there exists a one-to-one and onto function, $\mathcal{I} : V \to W$ such that

$$\mathcal{I}(\mathbf{x}^1 + \mathbf{x}^2) = \mathcal{I}(\mathbf{x}^1) + \mathcal{I}(\mathbf{x}^2)$$

for all $\mathbf{x}^1, \mathbf{x}^2 \in V$, and

$$\mathcal{I}(\alpha\mathbf{x}) = \alpha\mathcal{I}(\mathbf{x})$$

for all $\mathbf{x} \in V$, $\alpha \in \mathbb{F}$. Such an $\mathcal{I}$ is called an **isomorphism**.

The discussion preceeding this definition showed that all n-dimensional real[1] vector spaces are isomorphic to $\mathbb{R}^n$.

[1]Here "real" refers to the fact that the scalars for the vector space are real numbers. In this course, all vector spaces will be real. However, you can define vector spaces with scalars from other fields, such as the complex numbers.

2. Normed vector spaces

One property of two and three dimensional Euclidean space is that vectors have lengths. Our definition of vector spaces does not guarantee that we have a way of measuring length, so let's define a special type of vector space where we can measure length.

Definition 2.1. A **normed vector space**, $(V, +, \cdot, \|\cdot\|)$, is a vector space with a function, called the **norm**, from V to $\mathbb{R}$ and denoted by $\|v\|$ with the following properties:

(1) (Positive definite) $\|v\| \geq 0$ and $\|v\| = 0$ iff $v = 0$,
(2) (Homogenous) $\|\alpha v\| = |\alpha| \, \|v\|$ for all $\alpha \in \mathbb{R}$,
(3) The triangle inequality holds:
$$\|v_1 + v_2\| \leq \|v_1\| + \|v_2\|$$
for all $v_1, v_2 \in V$.

As above, when the addition, multiplication, and norm are clear from context, we will just write V instead of $(V, +, \cdot, \|\cdot\|)$ to denote a normed vector space. Like length, a norm is always non-negative and only zero for the zero vector. Also, similar to length, if we multiply a vector by a scalar, the norm also gets multiplied by the scalar. The triangular inequality means that norm obeys the idea that the shortest distance between two points is a straight line. If you go directly from x to y you "travel" $\|x - y\|$. If you stop at point z in between, you travel $\|x - z\| + \|z - y\|$. The triangle inequality guarantees that
$$\|x - y\| \leq \|x - z\| + \|z - y\|.$$
Any normed vector space is also a metric space with $d(x, y) = \|x - y\|$.

2.1. Examples.

Example 2.1. $\mathbb{R}^3$ is a normed vector space with norm
$$\|x\| = \sqrt{x_1^2 + x_2^2 + x_3^2}.$$
This norm is exactly how we usually measure distance. For this reason, it is called the Euclidean norm.

More generally, for any n, $\mathbb{R}^n$, is a normed vector space with norm
$$\|x\| = \sqrt{\sum_{i=1}^{n} x_i^2}.$$

The Euclidean norm is the most natural way of measuring distance in $\mathbb{R}^n$, but it is not the only one. A vector space can often be given more than one norm, as the following example shows.

Example 2.2. $\mathbb{R}^n$ with the norm
$$\|x\|_p = \left(\sum_{i=1}^{n} |x_i|^p \right)^{1/p}$$

for $p \in [1, \infty]$[a] is a normed vector space. This norm is called the p-norm.

[a]Where $\|x\|_\infty = \max_{1\le i\le n} |x_i|$

For nearly all practical purposes, $\mathbb{R}^n$ with any p-norm is essentially the same as $\mathbb{R}^n$ with any other p-norm. $\mathbb{R}^n$ is the same collection of elements regardless of the choice of p-norm, and the choice of p-norm does not affect the topology (i.e. which sets are open and closed) of $\mathbb{R}^n$ or the definition of derivatives. However, there are normed vector spaces where the choice of norm makes a difference.

Example 2.3. Let ℓ_p = infinite sequences such that

$$\|x\|_p = \left(\sum_{i=1}^{\infty} |x_i|^p\right)^{1/p}$$

is finite. Then ℓ_p with $\|\cdot\|_p$ is a normed vector space.

Example 2.4. Define

$$\|f\|_p = \left(\int_0^1 |f(x)|^p dx\right)^{1/p}.$$

Let $\mathcal{F}_p = \{f : (0,1)\to\mathbb{R} \text{ such that } \|f\|_p < \infty\}$. Define an equivalence relation between functions as

$$f \sim_p \tilde{f} \iff \|f - \tilde{f}\|_p = 0$$

The space $\mathcal{L}^p(0,1) = \{\text{equivalence classes of } \mathcal{F}_p\}$ with norm $\|\cdot\|_p$ is a normed vector space. The space is defined as the set of equivalences classes because if f and $\tilde{f}$ differ at only a finite collection of points[a], then

$$\|f - \tilde{f}\|_p = \left(\int_0^1 |f(x) - \tilde{f}(x)|^p dx\right)^{1/p} = 0.$$

If such f and $\tilde{f}$ were considered different vectors, then the norm would not be positive definite.

Moreover, $\mathcal{L}^p(0,1)$ is a different space for different p. For example, $\frac{1}{x^{1/p}} \notin \mathcal{L}^p(0,1)$, but $\frac{1}{x^{1/p}} \in \mathcal{L}^q(0,1)$ for $q < p$.

[a]Or, more generally, a set of measure zero.

Having a norm allows us to consider limits, continuity, and derivatives. When working with limits we will typically need to be sure that Cauchy sequences converge. In other words, we will want to work in complete normed vector spaces. Complete normed vector spaces are called **Banach** spaces.

3. Linear transformations

An isomorphism is a one-to-one and onto (bijective) functions that preserves addition and scalar multiplication. Can a function between vector spaces preserve addition and multiplication without being bijective? Let's try to construct an example. We know from above that all finite dimensional vector spaces are isomorphic to $\mathbb{R}^n$, so we might as well work with $\mathbb{R}^n$. To keep everything as simple as possible, let's just work with $\mathbb{R}^1$. Consider $f : \mathbb{R} \to \mathbb{R}$ defined by $f(x) = 0$ for all $x \in \mathbb{R}$. Clearly, f is not bijective. f preserves addition since

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

f also preserves multiplication because

$$\alpha f(x) = \alpha 0 = 0 = f(\alpha x).$$

Thus, we know there are functions that preserve addition and scalar multiplication but are not necessarily isomorphisms. Let's give such functions a name.

Definition 3.1. A **linear transformation** (aka linear function) is a function, A, from a vector space $(V, \mathbb{R}, +, \cdot)$ to a vector space $(W, \mathbb{R}, +, \cdot)$ such that $\forall v_1, v_2 \in V$,

$$A(v_1 + v_2) = Av_1 + Av_2$$

and

$$A(\alpha v_1) = \alpha A v_1$$

for all scalars $\alpha \in \mathbb{R}$.

A linear transformation from V to V is called a **linear operator** on V. A linear transformation from V to $\mathbb{R}$ is called a **linear functional** on V.

Any isomorphism between vector spaces is a linear transformation.

Example 3.1. Define $f : \mathbb{R}^2 \to \mathbb{R}$ by $f((x_1, x_2)) = x_1$, that is $f(x)$ is the first coordinate of x. Then,

$$f(\alpha x + y) = \alpha x_1 + y_1 = \alpha f(x) + f(y)$$

so f is a linear transformation.

In general we can construct linear transformations between finite dimensional vector spaces as follows. Let

$$A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

be a matrix. As usual let

$$A\mathbf{x} = \begin{pmatrix} \sum_{j=1}^n a_{1j} x_j \\ \vdots \\ \sum_{j=1}^n a_{mj} x_j \end{pmatrix},$$

for $\mathbf{x} = (x_1, ..., x_n) \in \mathbb{R}^n$. Then A is a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$. You may want to verify that $A(\alpha \mathbf{x}_1 + \mathbf{x}_2) = \alpha A \mathbf{x}_1 + A \mathbf{x}_2$ for scalars $\alpha \in \mathbb{R}$ and vectors $\mathbf{x}_1, \mathbf{x}_2 \in \mathbb{R}^n$.

Conversely let A be a linear transformation from V to W (if it is helpful, you can let $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$), and let $b_1, b_2, ..., b_n$ be a basis for V. By the definition of a basis, any $v \in V$ can be written $v = \sum_{j=1}^{n} v_j b_j$ for some $v_j \in \mathbb{R}$. By the definition of a linear transformation, we have

$$Av = \sum_{j=1}^{n} v_j A b_j.$$

Thus, a linear transformation is completely determined by its action on a basis. Also, if $d_1, \ldots, d_m$ is a basis for W then for each Ab_j we must be able to write Ab_j as a sum of the basis elements $d_1, \ldots, d_m$, i.e.

$$Ab_j = \sum_{i=1}^{m} a_{ij} d_i.$$

Substituting this equation into the previous one, we can write Av as

$$\begin{aligned} Av &= \sum_{j=1}^{n} v_j A b_j \\ &= \sum_{j=1}^{n} v_j \sum_{i=1}^{m} a_{ij} d_i \\ &= \sum_{i=1}^{m} d_i \left(\sum_{j=1}^{n} a_{ij} v_j \right) \end{aligned}$$

Thus, associated with a linear transformation there is an array of $a_{ij} \in \mathbb{R}$ determined by the linear transformation (and choice of basis for V and W). In the previous paragraph, we saw that conversely, if we have an array of $a_{ij} \in \mathbb{R}$ we can construct a linear transformation. This leads us to the following result.

Theorem 3.1. *For any linear transformation, A, from $\mathbb{R}^n$ to $\mathbb{R}^m$ there is an associated m by n matrix,*

$$\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

where a_{ij} is defined by $Ae_j = \sum_{i=1}^{m} a_{ij} e_i$. Conversely, for any m by n matrix, there is an associated linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ defined by $Ae_j = \sum_{i=1}^{n} a_{ij} e_i$.

Thus, we see that matrices and linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ are the same thing. This fact will help us make sense of many of the properties of matrices that we will go through in the next section. Also, it will turn out that most of the properties of matrices are properties of linear transformations. There are linear transformations that cannot be represented by matrices, yet many of the results and definitions that are typically stated for matrices will apply to these sorts of linear transformations as well.

Two examples of linear transformations that cannot be represented by matrices are integral and differential operators,

Example 3.2 (Integral operator). Let $k(x,y)$ be a function from $(0,1)$ to $(0,1)$ such that $\int_0^1\int_0^1 k(x,y)^2dxdy$ is finite. Define $K : \mathcal{L}^2(0,1) \to \mathcal{L}^2(0,1)$ by

$$(Kf)(x) = \int_0^1 k(x,y)f(y)dy$$

Then K is a linear transformation because

$$\begin{aligned}(K(\alpha f + g))(x) =& \int_0^1 k(x,y)(\alpha f(y) + g(y))dy \\ =&\alpha \int_0^1 k(x,y)f(y)dy + \int_0^1 k(x,y)g(y)dy \\ =&\alpha(Kf)(x) + (Kg)(x)\end{aligned}$$

Example 3.3 (Conditional expectation). One special type of an integral operator that appears often in economics is the conditional expectation operator. Suppose X and Y are real valued random variables with joint pdf $f_{xy}(x,y)$ and marginal pdfs $f_x(x) = \int_{\mathbb{R}} f(x,y)dy$ and $f_y(y) = \int_{\mathbb{R}} f(x,y)dx$. Consider the vector spaces

$$V = \mathcal{L}^2(\mathbb{R}, f_y) = \{g : \mathbb{R}{\to}\mathbb{R} \text{ such that } \int_{\mathbb{R}} f_y(y)g(y)^2dy < \infty\}$$

and

$$W = \mathcal{L}^2(\mathbb{R}, f_x) = \{g : \mathbb{R}{\to}\mathbb{R} \text{ such that } \int_{\mathbb{R}} f_x(x)g(x)^2dx < \infty\}$$

V is the space of all functions of Y such that the variance of $g(Y)$ is finite. Similarly, W is the space of all functions of X such that the variance of $g(X)$ is finite. The conditional expectation operator is $\mathcal{E} : V{\to}W$ defined by

$$(\mathcal{E}g)(x) = E[g(Y)|X = x] = \int_{\mathbb{R}} \frac{f_{xy}(x,y)}{f_x(x)f_y(y)} g(y)f_y(y)dy.$$

The conditional expectation operator is an integral operator, so it is a linear transformation.

Example 3.4 (Differential operator). Let $C^\infty(0,1)$ be the set of all infinitely differentiable functions from $(0,1)$ to $\mathbb{R}$. $C^\infty(0,1)$ is a vector space. Let $D : C^\infty(0,1) \to C^\infty(0,1)$ be defined by

$$(Df)(x) = \frac{df}{dx}(x)$$

Then D is a linear transformation.

Integral and differential operators are very important when studying differential equations. They are also useful in many areas of econometrics and in dynamic programming. We already encountered some linear transformations on infinite dimensional spaces when studying optimal control.

4. Matrix operations and properties

Let A and B be linear transformations from $\mathbb{R}^m$ to $\mathbb{R}^n$ and let $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ and $\begin{pmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{pmatrix}$ be the associated matrices. Since the linear transformation A and the matrix $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ represent the same object, we will use A to denote both. From the previous section, we know that A and B are characterized by their action on the standard basis vectors in $\mathbb{R}^n$. In particular, $Ae_j = \sum_{i=1}^m a_{ij}e_i$ and $Be_j = \sum_{i=1}^m b_{ij}e_i$.

4.1. **Addition.** To define matrix addition, it makes sense to require $(A+B)x = Ax + Bx$. Then,

$$\begin{aligned}(A+B)e_j =& Ae_i + Be_j \\ =& \sum_{j=1}^m a_{ij}e_i + \sum_{j=1}^m b_{ij}e_i \\ =& \sum_{j=1}^m (a_{ij}+b_{ij})e_i,\end{aligned}$$

so the only way sensible way to define matrix addition is

$$A+B = \begin{pmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{pmatrix}$$

As an exercise, you might want to verify that matrix addition has the following properties:

(1) Associative: $A + (B + C) = (A + B) + C$,
(2) Commutative: $A + B = B + A$,
(3) Identity: $A + \mathbf{0} = A$, where $\mathbf{0}$ is an m by n matrix of zeros, and
(4) Invertible $A + (-A) = \mathbf{0}$ where $-A = \begin{pmatrix} -a_{11} & \cdots & -a_{1n} \\ \vdots & \ddots & \vdots \\ -a_{m1} & \cdots & -a_{mn} \end{pmatrix}$.

4.2. **Scalar multiplication.** The definition of linear transformations requires that $A\alpha x = \alpha Ax$ where $\alpha \in \mathbb{R}$ and $x \in V$. To be consistent with this, for matrices we must define

$$\alpha A = \begin{pmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{pmatrix}$$

We have now defined addition and scalar multiplication for matrices. It should be no surprise that the set of all m by n matrices along with these two operations and the field $\mathbb{R}$ forms a vector space.

Example 4.1. The set of all m by n matrices is a vector space.

In fact, the above is not only true of the set of all m by n matrices, but of any set of all linear transformations between two vector spaces.

Example 4.2. Let $L(V, W)$ be the set of all linear transformations from V to W. Define addition and scalar multiplication as above. Then $L(V, W)$ is a vector space.

$L(\mathbb{R}^n, \mathbb{R}^m)$ is the set of all linear transformations from $\mathbb{R}^n \to \mathbb{R}^m$, i.e. all m by n matrices.

4.3. **Matrix multiplication.** Matrix multiplication is really the composition of two linear transformations. Let A be a linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$ and B be a linear transformation from $\mathbb{R}^p$ to $\mathbb{R}^n$. Now, we defined matrices by looking at how a linear tranformation acts on a basis vectors, so to define multiplication, we should look at $A(B\mathbf{g}_k)$. In these calculations, $\mathbf{e}_i$ are standard basis vectors in $\mathbb{R}^m$, $\mathbf{f}_j$ will be standard basis vectors in $\mathbb{R}^n$, and $\mathbf{g}_k$ will be basis vectors in $\mathbb{R}^p$.

$$
\begin{aligned}
A(B\mathbf{g}_k) &= A(\sum_{j=1}^{n} b_{jk}\mathbf{f}_j) && \text{definition of } B\mathbf{g}_k \\
&= \sum_{j=1}^{n} b_{jk}A\mathbf{f}_j && \text{Definition of linear transformtion} \\
&= \sum_{j=1}^{n} b_{jk}\left(\sum_{i=1}^{m} a_{ij}\mathbf{e}_i\right) && \text{definition of } A\mathbf{f}_j \\
&= \sum_{i=1}^{m}\left(\sum_{j=1}^{n} a_{ij}b_{jk}\right)\mathbf{e}_i \\
&= \begin{pmatrix} \sum_{j=1}^{n} a_{1j}b_{j1} & \cdots & \sum_{j=1}^{n} a_{1j}b_{jp} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{n} a_{mj}b_{j1} & \cdots & \sum_{j=1}^{n} a_{mj}b_{jp} \end{pmatrix}\mathbf{g}_k \\
&= (AB)\mathbf{g}_k.
\end{aligned}
$$

The indexing in the above equations is unpleasant and could be confusing. The important thing to remember is that matrix multiplication is the composition of linear transformations. It then makes sense that if A is m by n (a transformation from $\mathbb{R}^n$ to $\mathbb{R}^m$)and B is k by l (a transformation from R^l to $\mathbb{R}^k$), we can only multiply A times B if $k = m$. Matrix multiplication has the following properties:

(1) Associative: $A(BC) = (AB)C$
(2) Distributive: $A(B + C) = AB + AC$ and $(A + B)C = AC + BC$.

(3) Identity: $AI_n = A$ where A is m by n and I_n is the linear transformation from $\mathbb{R}^n$ to $\mathbb{R}^n$ such that $I_n x = x \forall x \in \mathbb{R}^n$.

Matrix multiplication is not commutative.

5. Null Spaces and Ranges

We are often interested in solving linear equations of the form $Ax = b$, where $x \in V$, $b \in W$, and $A \in L(V, W)$. For example, V could be $\mathbb{R}^n$, and W could be $\mathbb{R}^m$ and then A would be an $m \times n$ matrix. The null space and range of a linear transformation are two subspaces that can describe when the solution to $Ax = b$ is unique and when it the solution exists.

Definition 5.1. Let $A \in L(V, W)$. The set of solutions to the homogeneous equation $Ax = 0$ is the **null space** (or kernel) of A, denoted by $\mathcal{N}(A)$ (or null$A$),

$$\mathcal{N}(A) = \{x \in V : Ax = 0\}$$

As its name suggests, the null space of a linear transformation is a subspace.

Exercise 5.1. Show that $\mathcal{N}(A)$ is a linear subspace.

Null spaces are important for studying linear equations because if $z \in \mathcal{N}(A)$, then $A(x+z) = Ax + Az = Ax$. In other words, if $Ax = b$ for some x, then $A(x + z) = b$ for all $z \in \mathcal{N}(A)$. Note that $0 \in \mathcal{N}(A)$ always. From this discussion, we see that if $Ax = b$ has at least one solution, then it will have multiple solutions if $\mathcal{N}(A) \neq \{0\}$.

Definition 5.2. Let $A \in L(V, W)$. A is **one-to-one** (or **injective**) if $Ax = Av \implies x = v$.

Note that A is injective if and only if $\mathcal{N}(A) = 0$. Thus, we can also say that if $Ax = b$ has one solution, then it will have multiple solutions if A is not injective.

Definition 5.3. Let $A \in L(V, W)$. The **range** of A is the subset of W consisting of Ax for some $x \in V$, i.e.

$$\text{range}A = \{Ax : x \in V\} \subseteq W$$

When A is a matrix, its range is called its column space.

Exercise 5.2. Show that the range of a linear transformation is a linear subspace.

In terms of linear equations, $Ax = b$ has a solution if and only if $b \in \text{range}A$.

Definition 5.4. Let $A \in L(V, W)$, A is **onto** (or **surjective**) if range$A = W$.

The dimensions of the range and null spaces of a linear transformation are related.

Theorem 5.1 (Rank-Nullity theorem)**.** *If V is finite dimensional and $A \in L(V, W)$, then*

$$\dim(V) = \dim(\text{null}A) + \dim(\text{range}A).$$

Proof. Since V is finite dimensional and null$A \subseteq V$, null$A$ is also finite dimensional. Let $u_1, ..., u_n$ be a basis for null$A$. By lemma 1.3 we can expand this to a basis for V. Let,

$u_1, \dots, u_n, e_1, \dots, e_m$ be a basis for V, so $\dim(V) = n + m$. Let $v \in V$, then $\exists \alpha$'s and β's $\in \mathbb{R}$ such that

$$v = \sum_{i=1}^{n} \alpha_i u_i + \sum_{i=1}^{m} \beta_i e_i.$$

Since A is linear,

$$Av = \sum_{i=1}^{n} \alpha_i A u_i + \sum_{i=1}^{m} \beta_i A e_i.$$

Since $u_i \in \text{null} A$,

$$Av = \sum_{i=1}^{m} \beta_i A e_i.$$

Therefore, $Ae_1, \dots, Ae_m$ span range$A$.

We now show that $Ae_1, \dots, Ae_m$ must also be linearly independent. Suppose

$$\sum_{i=1}^{m} c_i A e_i = 0$$

then

$$A\left(\sum_{i=1}^{m} c_i e_i\right) = 0$$

so $\left(\sum_{i=1}^{m} c_i e_i\right) \in \text{null} A$. However, the u's span null$A$, so $\exists d$'s $\in \mathbb{R}$ such that

$$\sum_{i=1}^{n} d_i u_i = \sum_{i=1}^{m} c_i e_i.$$

Finally, since $u_1, \dots, u_n, e_1, \dots, e_m$ are linearly independent, the previous equation can only hold if the c's and d's are all 0. □

This theorem has some important implications for when a linear transformation can be one-to-one and onto.

Corollary 5.1. *If V is finite dimensional and* $\dim(V) > \dim(W)$*, then no linear transformation from V to W is one-to-one.*

Proof. Let $A \in L(V, W)$. A is one-to-one iff null$A = \{0\}$, i.e. iff dim(null A) = 0. From the theorem, this is impossible since

$$\dim(\text{null} A) = \dim(V) - \dim(\text{range} A) \geq \dim(V) - \dim(W) > 0$$

□

Corollary 5.2. *If W is finite dimensional and* $\dim(V) < \dim(W)$*, then no linear transformation from V to W is onto.*

Proof. Left as an exercise. □

5.1. **Norm for $L(V, W)$.** If V and W are normed vector spaces, then the space of linear transformations can also be given a norm.

Definition 5.5. A linear transformation $A : V \to W$ is bounded if there exists $M \in \mathbb{R}$ such that $\|Ax\|_W \leq M \|x\|_V$ for all $x \in V$.

Lemma 5.1. *A linear transformation is bounded if and only if it is continuous.*

Proof. On problem set 4. □

Lemma 5.2. *If V and W are finite dimensional, and $A \in L(V, W)$, then A is bounded.*

Proof. On problem set 4. □

In infinite dimensional spaces, there are discontinuous linear transformations. Let $B(V, W)$ denote the set of all bounded linear transformations from V to W.

Exercise 5.3. Show $B(V, W)$ is a linear subspace of $L(V, W)$.

$B(V, W)$ can be given norm:

$$\|A\|_{B(V,W)} = \sup_{x \neq 0, x \in V} \frac{\|Ax\|_W}{\|x\|_V}.$$

6. Transpose and dual spaces

Even more can be said about linear equations after we have defined the transpose of a linear transformation. Defining the transpose requires first introducing dual spaces.

Definition 6.1. Let V be a vector space. The **dual space** of V, denote V^* is the set of all (continuous)[2] linear functionals, $v^* : V \to \mathbb{R}$.

Example 6.1. The dual space of $\mathbb{R}^n$ is the set of $1 \times n$ matrices. In fact, for any finite dimensional vector space, the dual space is the set of row vectors from that space.

In fact, since any n dimensional vector space is isomorphic to $\mathbb{R}^n$, the dual space of any n dimensional space is the space itself. Dual spaces are especially important in economics because prices are in dual spaces.

Example 6.2 (Prices as elements of a dual space)**.** Suppose we have an economy with bundles of represented by vectors in some vector space, V. There could be n goods, and V could be $\mathbb{R}^n$. We could also think of the bundles of good as something like consumption at every instance of time and every state of the world. In that case, V would be a vector space of sequences (if time and states of the world are discrete) or functions (if time and states of the world are continuous). The dual space of V, V^* is

[2]All linear functionals on finite dimensional spaces are continuous. Some linear functionals on infinite dimensional spaces are not continuous. Depending on the text, the definition of dual space does not always require continuity. Sometimes the dual space is defined as the set of all linear functionals, and the topological dual space is the set of all continuous linear functionals.

the set of linear transformations from V to $\mathbb{R}$. If $p \in V^*$, then pv is the total cost of purchasing v. p is the price vector.

Dual spaces are also important in optimization because Lagrange multipliers are elements of a dual space.

Finite dimensional spaces are self-dual in the sense of example 6.1, i.e. V and V^* are isomorphic. Infinite dimensional spaces are often not self-dual, as in the following example.

Example 6.3. The space ℓ_p for $1 \leq p \leq \infty$ is the set of sequences of real numbers $\mathbf{x} = (x_1, x_2, ...)$ such that $\sum_{i=1}^{\infty} |x_i|^p < \infty$. (When $p = \infty$, $\ell_\infty = \{(x_1, x_2, ...) : \max_{i \in \mathbb{N}} |x_i| < \infty\}$). Such spaces appear in economics in discrete time, infinite horizon optimization problems.

Let's consider the dual space of ℓ_∞. In macro models, we rule out everlasting bubbles and ponzi schemes by requiring consumption divided by productivity to be in ℓ_∞. Every sequence, $\mathbf{p} = (p_1, p_2, ...) \in \ell_1$ gives rise to a linear functional on ℓ_∞ defined by

$$\mathbf{p}^*\mathbf{x} = \sum_{i=1}^{\infty} p_i x_i \leq \left(\sum_{i=1}^{\infty} |p_i|\right)\left(\max_{i \in \mathbb{N}} |x_i| < \infty\right).$$

We can conclude that $\ell_1 \subseteq \ell_\infty^*$.

As a (difficult) exercise, you could try to show whether or not $\ell_1 = \ell_\infty^*$. Exercise 3.46 of Carter is very related.

It is always however always the case that $V \subseteq (V^*)^*$.

Example 6.4. What is the dual space of $V = \mathcal{L}^2(\mathbb{R}, f_x) = \{g : \mathbb{R} \to \mathbb{R} \text{ such that } \int_{\mathbb{R}} f_x(x) g(x)^2 dx < \infty\}$? Let $h \in \mathcal{L}^2(\mathbb{R}, f_x)$. Define

$$h^*(g) = \int_{\mathbb{R}} f_x(x) g(x) h(x) dx.$$

Assuming $h^*(g)$ exists, h^* is an integral operator from V to $\mathbb{R}$, so it is linear. To show that $h^* \in V^*$ all we need to do is establish that $h^*(g)$ exists (is finite) for all $g \in V$. Hölder's inequality[a] , which we have not studied but is good to be aware of, says that

$$\int_{\mathbb{R}} f_x(x)|g(x)h(x)| dx \leq \sqrt{\int\int f_x(x) g(x)^2 dx}\sqrt{\int\int f_x(x) h(x)^2 dx}.$$

Since h and $g \in V$, the right hand side must be finite, so $h^*(g)$ is finite as well. Thus all such h^* is a subset of V^*. In fact, all such h^* is equal to V^*.[b]

We were actually working with V^* and similar dual spaces when we studied optimal control.

[a] See e.g. Wikipedia for a proof and more information.

[b] This is a consequence of the Riesz representation theorem.

Definition 6.2. If $A : V \to W$ is a linear transformation, then the **transpose** (or adjoint) of A is $A^T : W^* \to V^*$ defined by $(A^T w^*)v = w^*(Av)$.

To parse this definition, note that $A^T w^*$ is an element of V^*, so it is a linear transformation from V to $\mathbb{R}$. Thus, $(A^T w^*)v \in \mathbb{R}$. Similarly, $Av \in W$, and $w^* : W \to \mathbb{R}$, so $w^*(Av) \in \mathbb{R}$.

Example 6.5. Let $A \in L(\mathbb{R}^n, \mathbb{R}^m)$, so that A can be represented by an $m \times n$ matrix, $\begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$. Also, $A^T \in L(\mathbb{R}^m, \mathbb{R}^n)$ can be represented by an $n \times m$ matrix, $\begin{pmatrix} \tilde{a}_{11} & \cdots & \tilde{a}_{1n} \\ \vdots & \ddots & \vdots \\ \tilde{a}_{m1} & \cdots & \tilde{a}_{mn} \end{pmatrix}$. Let $v = e_k$ be the kth standard basis vector, and $w^* = e_j^*$. Then the definition of the transpose says that

$$
\begin{aligned}
(A^T e_j^*) e_k =& e_j^* A e_k \\
\left(\begin{pmatrix} \tilde{a}_{11} & \cdots & \tilde{a}_{1m} \\ \vdots & & \vdots \\ \tilde{a}_{n1} & \cdots & \tilde{a}_{nm} \end{pmatrix} e_j \right)^T e_k =& e_j^T \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix} e_k \\
\begin{pmatrix} \tilde{a}_{1j} & \cdots & \tilde{a}_{nj} \end{pmatrix} e_k =& e_j^T \begin{pmatrix} a_{1k} \\ \vdots \\ a_{mk} \end{pmatrix} \\
\tilde{a}_{kj} =& a_{jk}
\end{aligned}
$$

In other words, the definition is to simply swap rows and columns.

Exercise 6.1. What is the transpose of the conditional expectation operator from example 3.3?

Since the dual space of V is defined at $V^* = B(V, \mathbb{R})$, a norm on V^* can be defined in the same way,

$$\|v^*\|_{V^*} = \sup_{v \neq 0, v \in V} \frac{|v^* v|}{\|v\|_V}.$$

7. Separating hyperplane theorem

A line in $\mathbb{R}^2$ splits $\mathbb{R}^2$ into two pieces. A plane in $\mathbb{R}^3$ splits it into two pieces. More generally, an $n-1$ dimensional affine space splits $\mathbb{R}^n$ into two pieces.

Definition 7.1. A **hyperplane** in $\mathbb{R}^n$ is an $n-1$ dimensional affine subspace. Equivalently, a hyperplane is the set of solutions to a single equation with n variables.

Any hyperplane can be written in the form:

$$H_{\xi,c} = \{x : \xi^T x = c\}$$

where $c \in \mathbb{R}$ and $\xi \in (\mathbb{R}^n)^* = \mathbb{R}^n$. Based on this, we define a hyperplane in an arbitrary vector space as follows:

Definition 7.2. A **hyperplane** in V is any set that can be written as

$$H_{\xi,c} = \{x \in V : \xi x = c\}$$

for some $c \in \mathbb{R}$ and $\xi \in V^*$, $\xi \neq 0$

Hyperplanes play an important role in optimization. There is one theorem that is especially useful. We will use this theorem to prove the existence of Lagrange multipliers and the second welfare theorem. First, a definition.

Definition 7.3. A set $S \subseteq V$ is **convex** if $\forall x_1, x_2 \in S$ and $\lambda \in (0,1)$, we have $x_1\lambda + x_2(1-\lambda) \in S$.

If a set is convex, when we draw a line segment between any two points in the set, the line segment remains entirely within the set. In $\mathbb{R}^2$, convex sets include things shaped like triangles, squares, pentagons, circles, ellipses, etc. Some non-convex shapes are stars, horseshoes, rings, etc.

Theorem 7.1 (Separating hyperplane theorem). *If S_1 and $S_2 \subseteq V$ are convex and $S_1 \cap S_2 = \emptyset$ and either V is finite dimensional or the (algebraic) interior of S_1 or S_2 is not empty. Then there exists a hyperplane, $H_{\xi c} = \{x : \xi x = c\}$ such that*

$$\xi s_1 \leq c \leq \xi s_2$$

for all $s_1 \in S_1$ and $s_2 \in S_2$. We say that $H_{\xi,c}$ separates S_1 and S_2.

Visually, this theorem says that we can draw a hyperplane, H, between S_1 and S_2. H is orthogonal to the line passing through ξ and 0. The projection of S_1 on ξ is disjoint from the projection of S_2 on ξ. See figure 1 for an illustration in $\mathbb{R}^2$.

Exercises 3.182-3.186 of Carter (2001) guide you through a proof of the separating hyperplane theorem in $\mathbb{R}^n$. A proof of the general version can be found in appendix A, or any text on functional analysis such as Luenberger (1969), Clarke (2013), or Holmes (1975).

Comment 7.1. It is often useful to refine the separating hyperplane to obtain either:

- strict separation — one or both weak inequalities become strict, or
- supporting hyperplane — the separating hyperplane theorem holds as stated, and $\exists\, s_1^* \in S_1$ such that $\xi s_1^* = c$

Strict separation holds for all $s_1 \in \text{aint}(S_1)$ and $s_2 \in \text{aint}(S_2)$, $\xi s_1 < c < \xi s_2$. Thus, if you need strict separation of some S_1 and S_2, one technique is to show that there are disjoint convex sets A_1 and A_2 such that $S_1 \subseteq \text{int}(A_1)$ and $S_2 \subseteq \text{int}(A_2)$.

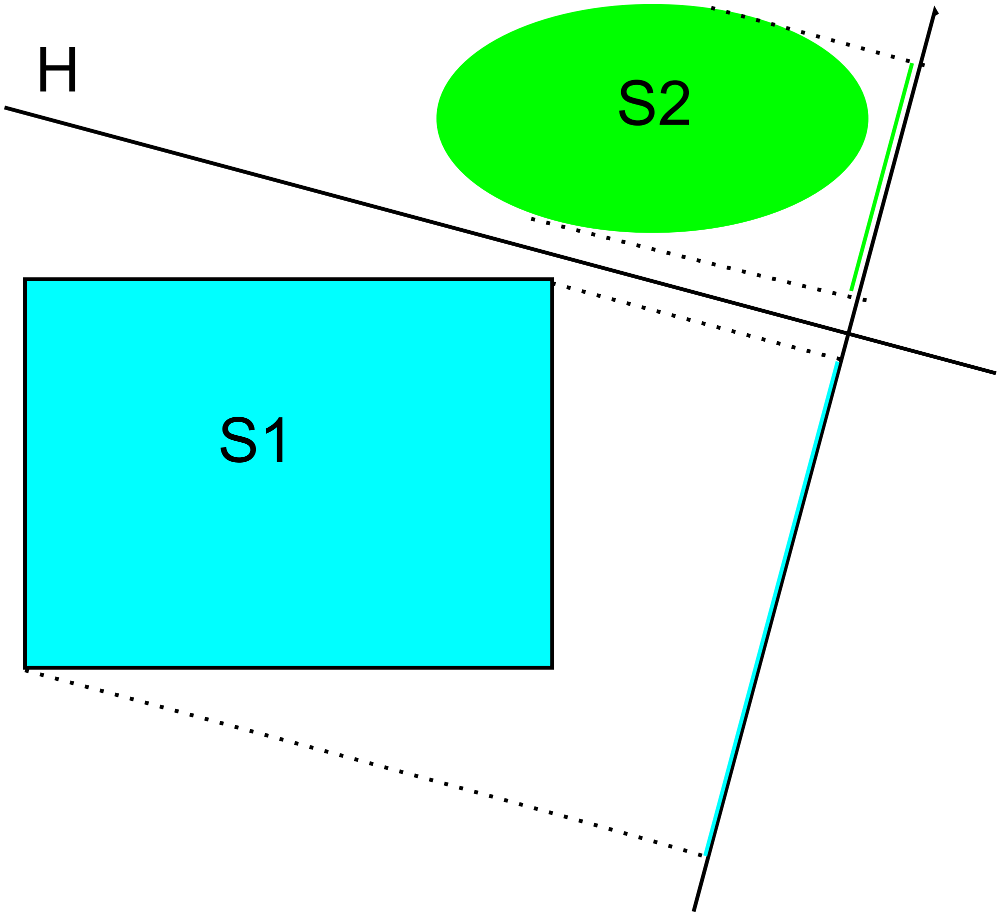
There exists a supporting hyperplane at s_1^* if s_1^* is in the (algebraic) boundary of S_1 and S_1 has a non-empty interior.

As a challenging exercise, you could try to prove the preceding statements.

7.1. **Existence of Lagrange multipliers.** In studying constrained maximization problems, we showed that if x^* is a local maximizer for

$$\max f(x) \text{ s.t. } h(x) = c$$

FIGURE 1. Separating hyperplane



then it must be that for all v such that $Dh_{x^*}v = 0$ we also have $Df_{x^*}v = 0$. We then made some heuristic arguments that this is equivalent to $Df_x = \mu^T Dh_x$ for some Lagrange multipliers μ. The separating hyperplane theorem let's us prove this fact. Notice that Df_{x^*} is a $1 \times n$ matrix, i.e. a linear transformation from $\mathbb{R}^n \to \mathbb{R}$, and Dh_{x^*} is an $m \times n$ matrix, i.e. a linear transformation from $R^n \to \mathbb{R}^m$.

Theorem 7.2. *Let V and W be normed vector spaces, $A \in B(V, \mathbb{R})$ and $C \in B(V, W)$. Assume that* range$C$ *is closed. Then* null$A \supseteq$ null$C$ *if and only if $A = \mu C$ for some $\mu \in W^*$.*

Proof. Assume null$A \supseteq$ null$C$. Consider $D : V \to \mathbb{R} \times W$ defined by $Dv = (Av, -Cv)$. D is linear and bounded since A and C are. Therefore range$D$ is a subspace of $\mathbb{R} \times W$. Also $(1, 0_W) \notin$ range$D$ because $Cv = 0_w$ only if $v \in$ null$C$, but if $v \in$ null$C$, then $v \in$ null$A$ by assumption, and $Av = 0 \neq 1$. Since range$C$ is closed and range$A$ is finite dimensional and therefore closed, range$D$ is closed. Therefore, there exists an open neighborhood of $(1, 0_W)$ that does not intersect range$D$. Call it N. This neighborhood is convex, and so is range$D$. Therefore, by the separating hyperplane theorem, there exists $\xi \in (\mathbb{R} \times W)^*$ and $c \in \mathbb{R}$ such that

$$\xi x \leq c < \xi y$$

for all $x \in \text{range}D$ and $y \in N$. It must be that $\xi x = 0$ for all $x \in \text{range}D$. If not, say $\xi x = d \neq 0$, then since range$D$ is a subspace, $(2c/d)x \in \text{range}D$, and since ξ is linear, $\xi((2c/d)x) = 2c > c$. Let $\xi = (\xi_1, -\tilde{\mu})$, where $\xi_1 \in \mathbb{R}^* = \mathbb{R}$ and $\mu \in W^*$. Since $(1, 0_w) \in N$, it must be that $\xi_1 > 0$. Therefore, for all $v \in V$,

$$\begin{aligned} 0 =& \xi D v \\ 0 =& \xi_1 A v - \tilde{\mu} C v \\ \mu C v =& A v \end{aligned}$$

i.e. $\mu C = A$.

Conversely, suppose $\mu C = A$. Let $v \in \text{null}C$. Then, $0 = Cv$, so $Av = \mu C v = \mu 0 = 0$. Therefore $v \in \text{null}A$. □

If we take $A = Df_x$ and $C = Dh_x$, then this theorem shows that existence of Lagrange multipliers such that $Df_x = \mu^T Dh_x$ is equivalent to $Dh_{x^*} v = 0$ implying that $Df_{x^*} v = 0$.

8. Welfare theorems

A second major use of the separating hyperplane theorem (and vector spaces more generally) is in the proof of the first and second welfare theorems. The first welfare theorem says that every competitive equilibrium is Pareto efficient. The second welfare theorem says that every Pareto efficient allocation can be achieved by some competitive equilibrium.

We have some set of commodities, S, which we will assume is a normed vector space. For example, in a world with n goods, S could be $\mathbb{R}^n$ and for each $s = (s_1, ..., s_n) \in S$, s_j represents the quantity of the jth good. These goods include everything that is bought or sold, including things like food or clothing that we usually think of as goods, and things like labor and land. There are I consumers, indexed by i. Each consumer chooses goods from a feasible set $X_i \subseteq S$. These X_i are feasible consumption sets, not budget sets. It is supposed to represent the physical constraints of the world. For example if there are three goods: food, clothing, and labor measured in days of labor per day, then X_i might be $[0, \infty) \times [0, \infty) \times [0, 1]$. Each consumer has preferences over X_i represented by a preference relation, $\succeq_i$ that as in the previous lecture the following properties:

(1) (complete) $\forall x, z \in X_i$, either $x \succeq_i z$ or $z \succeq_i x$ or both,
(2) (transitive) $\forall x, w, z \in X_i$, if $x \succeq_i w$ and $w \succeq_i z$ then $x \succeq_i z$,
(3) (reflexive) $\forall x \in X_i$, $x \succeq_i x$.

In words, $x \succeq_i z$ means person i likes the bundle of goods x as much as or more than the bundle of goods z. If you wish, you can think of the preference relation coming from a utility function, $u_i(x) : X_i \to \mathbb{R}$ and $x \succeq_i z$ means $u_i(x) \geq u_i(z)$. If $x \succeq_i z$ but $z \not\succeq_i x$, then we say that x is strictly preferred to z and write $x \succ_i z$. If $x \succeq_i z$ and $z \succeq_i x$ we say that person i is indifferent between x and z and write $x \simeq_i z$.

There are also J firms indexed by j. Each firm j chooses production y_j from production possibility set $Y_j \subseteq S$. The firm will produce positive quantities of its outputs and negative quantities of its inputs. Continuing with the example of three goods, if the firm produces

$F^f(l)$ units of food from l units of labor and $F^c(l)$ units of clothing, then production possibility set could be written:

$$Y_j = \{(f, c, l) \in S : l \leq 0 \wedge f \leq F^f(\alpha|l|) \wedge c \leq F^c((1-\alpha)|l|) \text{ for some } \alpha \in [0, 1]\}.$$

Firms produce goods and consumers consume goods. For the market to **clear** we must have sum of production equal to the sum of consumption, i.e.

$$\sum_{i=1}^{I} x_i = \sum_{j=1}^{J} y_j$$

We call the $I + J$-tuple of all x_i and y_j, $\big((x_1, ..., x_I), (y_1, ..., y_J)\big)$ (which we will sometimes shorted by just writing $((x_i), (y_j)))$ an **allocation**. An allocation is **feasible** if $x_i \in X_i \forall i$, $y_j \in Y_j \forall j$, and $\sum_{i=1}^{I} x_i = \sum_{j=1}^{J} y_j$.

Definition 8.1. An allocation, $((x_i^0), (y_j^0))$, is **Pareto efficient** (or Pareto optimal) if it is a feasible and there is no other feasible allocation, $((x_i), (y_j))$, such that $x_i \succeq_i x_i^0$ for all i and $x_i \succ_i x_i^0$ for some i.

This definition is just a mathematical way of stating the usual verbal definition of Pareto efficient. An allocation is Pareto efficient if there is no other allocation that makes at least one person better off and no one worse off.

We are going to be comparing competitive equilibria to Pareto efficient allocations. To do that we must first define a competitive equilibrium. A price system is a continuous linear transformation, $p : S \to \mathbb{R}$, i.e. $p \in S^*$. In the case where $S = \mathbb{R}^n$, a price system is just a $1 \times n$ matrix. The entries in this price matrix are the prices of each of the n goods. px for $x \in S$ represents the total expenditure needed to purchase the bundle of goods x.

Definition 8.2. An allocation, $((x_i^0), (y_j^0))$, along with a price system, p, is a **competitive equilibrium** if

(C1) The allocation is feasible
(C2) For each i and $x \in X_i$ if $px \leq px_i^0$ then $x_i^0 \succeq_i x$,
(C3) For each j if $y \in Y_j$ then $py \leq py_j^0$

Condition C2 says that each consumer must be choosing the most preferred bundle of goods that he or she can afford. If the preference relation comes from a utility function, C2 says that consumers maximize their utility given prices. Similarly, condition C3 says that producers maximize profits.

The first welfare theorem requires one additional condition on preferences.

Definition 8.3. Preference relation $\succ_i$ has the **local non-satiation condition** if for each $x \in X_i$ and $\epsilon > 0$ $\exists x' \in X_i$ such that $\|x - x'\| \leq \epsilon$ and $x' \succ_i x$.

This condition says that given any bundle of goods you can find another bundle very close by that is preferred. If the preference relation comes from utility function, the utility function having a non-zero derivative everywhere implies local non-satiation. The intuition for why the first welfare theorem requires local non-satiation is that local non-satiation rules out the following scenario. Suppose person i does not care about clothing at all. Then you take clothes away from person i, making person i no worse off, and give them to someone else, making that person better off. However, there is nothing in the definition of a competitive equilibrium that prevents person i from having clothes.

8.1. **First welfare theorem.**

Theorem 8.1 (First welfare theorem)**.** *If* $((x_i^0), (y_j^0))$ *and* p *is a competitive equilibrium and all consumers' preferences have the local non-satiation condition, then* $((x_i^0), (y_i^0))$ *is Pareto efficient.*

Proof. We will prove it by contradiction. Suppose that a competitive equilibrium is not Pareto efficient. Then there exists another feasible allocation[3], $((x_i), (y_j))$, such that there is at least one $x_{i^*} \succ_{i^*} x_{i^*}^0$. The contrapositive of condition C2 in the definition of competitive equilibrium implies that then $px_{i^*} > px_{i^*}^0$. For all other $i \neq i^*$ it must be that $x_i \succeq_i x_i^0$. When $x_i \succ_i x_i^0$, by the same argument as above, $px_i > px_i^0$. When $x_i \simeq_i x_i^0$, then we will show that local non-satiation implies $px_i \geq px_i^0$. If not and $px_i < px_i^0$, then by continuity of p there exists some $\delta > 0$ such that for all x' with $\|x_i - x'\| < \delta$, we have

$$|px_i - px'| < |px_i - px_i^0|$$

and in particular,

$$px' < px_i^0.$$

Additionally since preferences are locally non-satiated, there exists some $\tilde{x}$ with $\|x_i - \tilde{x}\| < \delta$ and $\tilde{x} \succ_i x_i \simeq_i x_i^0$. However, then we also have $\tilde{x} \succ_i x_i^0$ and $p\tilde{x} < px_i^0$, which contradicts x_i^0 and p being part of a competitive equilibrium. Thus, we can conclude that $px_i \geq px_i^0$.

At this point we have shown that if $((x_i^0), (y_j^0))$ is a competitive equilibrium that is not Pareto efficient, then there is some other allocation $((x_i), (y_j))$ that is feasible and has $x_i \succeq_i x_i^0$, which implies that $px_i \geq px_i^0$. Each consumer spends (weakly) more in this alternative, Pareto improving allocation. Now we will show that each consumer spending at least as much contradicts profit maximization. The total expenditure of consumers in the alternate allocation must be greater than in the competitive equilibrium because there is one consumer who is spending strictly more. That is,

$$\sum_{i=1}^{I} px_i > \sum_{i=1}^{I} px_i^0 \tag{1}$$

The price system is a linear transformation, so

$$\sum_{i=1}^{I} px_i = p\left(\sum_{i=1}^{I} x_I\right)$$

[3]This sort of allocation is called a Pareto improvement.

Both allocations are feasible, and, in particular, market clearing so

$$\sum_{i=1}^{I} x_i = \sum_{j=1}^{J} y_j$$

Applying p to both sides,

$$p\left(\sum_{i=1}^{I} x_i\right) = p\left(\sum_{j=1}^{J} y_j\right)$$
$$= \sum_{j=1}^{J} p y_j.$$

Identical reasoning would show that

$$\sum_{i=1}^{I} p x_i^0 = \sum_{j=1}^{J} p y_j^0.$$

Substituting into (1) we get

$$\sum_{j=1}^{J} p y_j > \sum_{j=1}^{J} p y_j^0. \tag{2}$$

But this contradicts profit maximization (C3) since $y_j \in Y_j$ and we cannot have (2) if $p y_j \leq p y_j^0$. Therefore, we conclude that there can be no Pareto improvement from a competitive equilibrium, i.e. any competitive equilibrium is Pareto efficient. □

8.2. **Second welfare theorem.** The second welfare theorem is the converse of the first welfare theorem. The second welfare theorem says that any Pareto efficient allocation can be achieved by some competitive equilibrium. The second welfare theorem does not hold quite as generally as the first welfare theorem.

Definition 8.4. A preference relation, $\succeq_i$, is **convex** if whenever $x \succeq_i z$ and $y \succeq_i z$, then $\lambda x + (1-\lambda) y \succeq_i z$ for all $\lambda \in [0,1]$.

Alternatively, a preference relation is convex if the set $\{x \in X_i : x \succeq_i z\}$ is convex for each z. Whenever you have seen convex indifference curves, the associated preference relation is convex. If the preference relation is generated by a concave (more generally quasi-concave) utility function, then the preference relation is convex.

Definition 8.5. A preference relation, $\succeq_i$, is **continuous** if for any $x \succ_i z$ there exists a $\delta > 0$ such that for all x' with $\|x - x'\| < \delta$ we have $x' \succ_i z$.

A continuous preference relation can be generated by a continuous utility function.

Theorem 8.2 (Second welfare theorem)**.** *Assume the preferences of each consumer are convex, locally non-satiated, and continuous, and that X_i is convex and non-empty. Also assume that Y_j is convex and non-empty for each firm j.*

Suppose $((x_i^e),(y_j^e))$ is a Pareto efficient allocation such that for any price system, p, there is always a cheaper bundle of goods, i.e. $\exists x_i \in X_i$ s.t. $px_i < px_i^e$ for each i. Then there exists a price system, p^e such that $((x_i^e),(y_j^e))$ and p^e is a competitive equilibrium.

Proof. We are going to construct the price system by applying the separating hyperplane theorem. Let $V_i = \{x \in X_i : x \succ_i x_i^e\}$ be the set of x strictly preferred by person i. Let

$$V = \{\chi \in S : \chi = \sum_{i=1}^{I} x_i \text{ where } x_i \in V_i\}$$

be the set of sums of elements from each V_i. The convexity of X_i and the preference relation implies that V_i is convex for each i. That, in turn, implies that V is convex.[4] Similarly, if

$$Y = \{\psi \in S : \psi = \sum_{i=j}^{J} y_j \text{ where } y_j \in Y_j\}$$

is the sum of each firms' production possibility set, then Y is convex.

We have two convex sets. Now, we just need to show that they are disjoint, and then we can apply the separating hyperplane theorem. Suppose $\chi \in Y \cap V$. Then $\exists x_i \in V_i$ and $y_j \in Y_j$ such that $\chi = \sum_{i=1}^{I} x_i = \sum_{j=1}^{J}$. This is feasible allocation, and $x_i \succ_i x_i^e$ by construction. This contradicts $((x_i^e),(y_i^e))$ being Pareto efficient. Therefore, $Y \cap V = \emptyset$. o

Now, by the separating hyperplane theorem, $\exists p \in S^*$ and $c \in \mathbb{R}$ such that[5]

$$p\chi \geq c \geq p\psi \tag{3}$$

for all $\chi \in V$ and $\psi \in Y$. Now we need to verify that $((x_i^e),(y_j^e))$ with p is a competitive equilibrium. It is feasible because $((x_i^e),(y_j^e))$ is Pareto efficient, and feasible by definition.

We now show that (3) holds with $c = p\chi^e = p\psi^e$, where $\chi^e = \sum_{i=1}^{I} x_i^e$ and $\psi^e = \sum_{j=1}^{J} y_j^e$. On the one hand, $\chi^e = \psi^e \in Y$, so we must have

$$c \geq p\chi^e$$

On the other hand, for any $\delta > 0$, by local non-satiation, we can find x_i such that $x_i \succ_i x_i^e$ and $\left\|x_i - x_i^e\right\| < \delta/I$. It follows from the triangle inequality that $\left\|\sum_{i=1}^{I} x_i - \sum_{i=1}^{I} x_i^e\right\| < \delta$. p is continuous, so for any $\epsilon > 0$ we can find a δ small enough that

$$\left| p\left(\sum_{i=1}^{I} x_i^e\right) - p\left(\sum_{i=1}^{I} x_i\right) \right| < \epsilon,$$

[4]It might be a good exercise to prove these claims.

[5]In the notation of theorem 7.1, p is ξ.

Then, for any $\epsilon > 0$, there exists $x_i \in V_i$ such that

$$p\chi^e + \epsilon > p(\sum x_i) \geq c$$

Since this is true for any ϵ, it must be that $p\chi^e \geq c$. Therefore, we have now shown that

$$p\chi \geq c = p\chi^e = p\psi^e \geq p\psi \tag{4}$$

for all $\chi \in V$ and $\psi \in Y$.

We now show that firms and consumers are maximizing given this price. Let $y_\ell \in Y_\ell$. Then $\sum_{j \neq \ell} y_j^e + y_\ell \in Y$, so

$$p\left(\sum_{j\neq\ell} y_j^e + y_\ell\right) \leq p\psi^e = p\left(\sum_j y_j^e\right)$$
$$py_\ell \leq py_\ell^e$$

Thus, each firm is maximizing profits given p.

Now we show that consumers are maximizing. It must be then also be that $px_i \geq px_i^e$ for each i and all $x_i \in V_i$. If not, then there is an $\epsilon > 0$ such that $px_i + \epsilon < px_i^e$, and then using local non-satiation we can choose x_k for $k \neq i$ such that $x_k \in V_k$ and

$$\left|\sum_{k\neq i} px_k - \sum_{k\neq i} px_k^e\right| < \epsilon/2$$

and then

$$\sum_{k=1}^{I} px_k + \epsilon/2 < \sum_{k=1}^{I} px_k^e.$$

Similarly, we must have $py_j^e \geq py_j$ for all $y_j \in Y_j$, which proves that profit maximization, (C3), holds.

We have nearly shown that utility maximization, (C2), also holds. We have shown that for each i if $x_i \succ_i x_i^e$ then $px_i \geq px_i^e$. To strengthen it to the form in the definition, we need to show that $px_i > px_i^e$. We will use the continuity of preferences and the cheaper good condition. Suppose $px_i = px_i^e$ and $\exists x_i' \in X_i$ such that $px_i' < px_i^e$. Then for any $\lambda \in (0,1)$, $p(\lambda x_i' + (1-\lambda)x_i') < px_i^e$. Also, by the continuity of preferences, for λ close enough to 0, $\lambda x_i + (1-\lambda)x_i' \succ_i x_i^e$. However, then $\lambda x_i + (1-\lambda)x_i' \in V_i$ contradicting $p(\lambda x_i + (1-\lambda)x_i') < px_i^e$. Therefore, if the cheaper good exists, we must have $px_i < px_i^e$. □

Appendix A. Proof of the separating hyperplane theorem

This section is based on Holmes (1975). The following lemma is a first step to proving the theorem.

Lemma A.1 (Stone)**.** *Let V be a vector space and A and B be disjoint convex subsets of V, then there exist convex sets C and D such that $A \subseteq C$, $B \subseteq D$, $C \cap D = \emptyset$ and $C \cup D = V$.*

Proof. We employ a similar technique as used to show the existence of a basis. Let $\mathcal{C}$ be the collection of all convex sets containing A and disjoint from B. $\mathcal{C}$ is non-empty because $A \in \mathcal{C}$. As in the exercise showing existence of a basis, $(\mathcal{C}, \subseteq)$ is a partially ordered set. Also, for any chain $\mathcal{H} \subseteq \mathcal{C}$, $\cup_{E \in \mathcal{H}} E$ is an upper bound. Therefore, by Zorn's lemma, $\mathcal{C}$ has a maximal element. Let C be a maximal element. Define $D = V \setminus C$. By construction $C \cap D = \emptyset$, $B \subseteq D$, and $C \cup D = V$.

To complete the proof, we just need to show that D is convex. Suppose D is not convex, then there exists $x, z \in D$, $\lambda \in (0,1)$ such that $y = \lambda x + (1-\lambda)z \notin D$. Then $y \in C$. Furthermore, observe that for any $d \in D$, there must be a $c \in C$ and $\lambda \in [0,1]$ such that $b = \lambda d + (1-\lambda)c \in B$. If there were not, then the set

$$\tilde{C} = \{\lambda d + (1-\lambda)c : \lambda \in [0,1], c \in C\}$$

would be a convex set containing A and disjoint from B, contradicting the fact that C is a maximal such set. Then can find $p, q \in C$ and $u, v \in B$ such u is a convex combination of p and x, and v is a convex combination of q and z, i.e. there are $\lambda_u, \lambda_v \in (0,1)$ such that:

$$u = \lambda_u x + (1-\lambda_u)p$$
$$v = \lambda_v z + (1-\lambda_v)q$$

However, then there would be some convex combination of u and v that can also be written as a convex combination of p, q, and y. Since B is convex and $u, v \in B$, this convex combination would also be in B. Since C is convex and $p, q, y \in C$, the convex combination would also be in C. In other words $B \cap C$ would not be empty, but this contradicts the way C is defined. Therefore, it must be that D is convex. □

For any hyperplane, $H_{\xi,c}$, the sets $\{v \in V : \xi v \geq c\}$ and $\{v \in V : \xi v < c\}$ are disjoint and convex. The next step is to show that disjoint convex sets from lemma A.1 take this form. First, we need some definitions of the boundary and interior of subsets of a vector space (without a norm). For convex sets in finite dimensional normed vector spaces, these definitions are the same as the definitions using open and closed sets. In general, the definitions differ, but they are capturing similar ideas.

Definition A.1. Let V be a vector space and $A \subseteq V$. The **algebraic interior** of A is set of all $a \in A$ such that for every $v \in V$ there exists $\bar{\lambda} \in (0,1)$ such that

$$a(1-\lambda) + \lambda v \in A$$

for all $\lambda \in [0, \bar{\lambda}]$. Denote the algebraic interior of A as $\text{int}^{\mathcal{A}}(A)$.

At interior points, we can move slighty from a toward any other point v and remain inside A. This algebraic interior is different than the topological interior we defined using open sets earlier. For example, in ℓ^∞ let $A = \{(x_1, x_2, ...) : |x_n| < 1/n\}$. A is equal to its algebraic interior, but the topological interior of A is empty. As in this example, it is always the case that the topological interior is contained in the algebraic interior.

Lemma A.2. *Let V be a normed vector space, and $A \subseteq V$. Then the topotical interior of A is (weak) subset of the algebraic interior of A.*

Proof. Left as an exercise. □

In the statement of the separating hyperplane theorem, we said that "the interior of S_1 or S_2 is not empty." In this section, we will prove the theorem with the assumption that the *algebraic* interior of one the set is not empty. Since a non-empty topological interior implies a non-empty algebraic interior, the theorem is also true (but slightly less general) if we interpret "interior" as meaning topological interior instead.

Definition A.2. Let V be a vector space and $A \subseteq V$. The **linear closure**[6] of A is set of all $v \in V$ such that $\exists a \in A$ such that $\forall \lambda \in [0, 1)$,

$$a(1 - \lambda) + \lambda v \in A$$

Denote the linear closure of A as $\overline{A}^{\mathcal{A}}$.

Finally, we need to define affine sets.

Definition A.3. $A \subseteq V$ is **affine** if $\forall x, y \in A$ and $\lambda \in \mathbb{R}$, $\lambda x + (1 - \lambda)y \in A$.

The difference between affine and convex is that affine sets allow $\lambda < 0$ and $\lambda > 1$. Affine sets contain the line passing through any two vectors in the set. Convex sets only contain the line segment between any two points. Affine sets are like linear subspaces in that they are lines, planes, etc, except that affine sets need not contain 0. We also need the following lemma about hyperplanes.

Lemma A.3. *An affine set A is a hyperplane if and only if A is proper subset that is maximal with respect to inclusion.*

Proof. Left as an exercise. □

Lemma A.4. *Let C and D be non-empty convex sets in a vector space V with $C \cap D = \emptyset$ and $C \cup D = V$. Let $M = \overline{C}^{\mathcal{A}} \cap \overline{D}^{\mathcal{A}}$, then either $M = V$ or M is a hyperplane in V.*

Proof. Since C and D are convex, so are $\overline{C}^{\mathcal{A}}$ and $\overline{D}^{\mathcal{A}}$. For any convex sets, their intersection is also convex, so M is convex. Moreover, M is affine. To see this let $x, y \in M$ and $z = \lambda x + (1 - \lambda)y$ for $\lambda \in \mathbb{R}$. If $\lambda \in (0, 1)$, we known M is convex, so $z \in M$. If $z \notin M$, then $z \in \text{int}^{\mathcal{A}}(C) \cup \text{int}^{\mathcal{A}}(D)$. To be concrete, suppose $z \in \text{int}^{\mathcal{A}}(C)$ and $\lambda < 1$ (the other cases are dealt with similarly). Then $y = z/(1 - \lambda) - \lambda/(1 - \lambda)x$ is a convex combination of x and z. Notice that $x \in \overline{C}^{\mathcal{A}}$ and $z \in \text{int}^{\mathcal{A}}(C)$ implies $y \in \text{int}^{\mathcal{A}}(C)$, but this contradicts the assumption that $y \in M$. Therefore, it must be that $z \in M$.

[6]This is non-standard terminology, Holmes (1975) calls this lin(A). Holmes's notation for my $\text{int}^{\mathcal{A}}(A)$ is cor(A).

Suppose $M \neq V$. Pick any $m \in M$. Then $\exists (m+p) \in V \setminus M = \text{int}^{\mathcal{A}}(C) \cup \text{int}^{\mathcal{A}}(D)$. Without loss of generality, assume $(m+p) \in \text{int}^{\mathcal{A}}(C)$. If $(m-p) \in \text{int}^{\mathcal{A}}(C)$, then $0.5(m+p) + 0.5(m-p) = m \in \text{int}^{\mathcal{A}}(C)$ by convexity, but we know $m \notin \text{int}^{\mathcal{A}}(C)$, so it must be that $(m-p) \in \text{int}^{\mathcal{A}}(D)$. To conclude, we show that the span of $\{p\} \cup M = V$, so M is a maximal affine set, i.e. a hyperplane. Let $x \in C$. By definition of $\overline{C}^{\mathcal{A}}$ and $\overline{D}^{\mathcal{A}}$, there exists $\lambda \in (0,1)$ such that $\lambda x + (1-\lambda)(m-p) \in M$, so $x \in \text{span}(M \cup \{p\})$. Similarly if $x \in D$, $x \in \text{span}(M \cup \{p\})$. □

Finally, we can prove the separating hyperplane theorem.

Proof of theorem 7.1. By Stone's separation lemma A.1, there exists convex sets C and D such that $S_1 \subseteq C$, $S_2 \subseteq D$, $C \cap D = \emptyset$, and $C \cup D = V$. By lemma A.4, $H = \overline{C}^{\mathcal{A}} \cap \overline{D}^{\mathcal{A}}$ is either V or the separating hyperplane that we want. H is not a hyperplane only if $\overline{C}^{\mathcal{A}} = \overline{D}^{\mathcal{A}} = V$. If V is finite dimensional, and C and D are not empty, then this is impossible. Proving this is left as an exercise. Regardless of the dimension of V, if C (or D) has a non-empty interior, then lemma A.5 implies that $\overline{C}^{\mathcal{A}}$ (or $\overline{D}^{\mathcal{A}}$) cannot be all of V, since we know that $C \neq V$. □

Lemma A.5. *If $C \subsetneq V$ is convex,* $\text{int}^{\mathcal{A}}(C)$ *is not empty, and* $\overline{C}^{\mathcal{A}} = V$, *then* $C = V$.

Proof. Let $x \in V$, and $y \in \text{int}^{\mathcal{A}}(C)$. Without loss of generality assume $y = 0$. Since $\overline{C}^{\mathcal{A}} = V$ so $2x \in \overline{C}^{\mathcal{A}}$, $\exists z \in C$ such that

$$(1-\lambda)z + \lambda 2x \in C$$

for all $\lambda \in [0,1)$. Since $y = 0 \in \text{int}^{\mathcal{A}}(C)$ there is $\delta > 0$ such that $-\delta z \in C$. Since C in convex, for any $t \in [0,1]$ and $\lambda \in [0,1)$,

$$[(1-\lambda)z + \lambda 2x]t - (1-t)\delta z \in C$$

Setting $\lambda = \frac{1+\delta}{1+2\delta}$ and $t = \frac{1+2\delta}{2(1+\delta)}$, we can conclude that $x \in C$, so $C = V$. □

Appendix B. Fundamental theorem of linear algebra

I've covered this in past years, but we will not get to it this year. For the purposes of this course, the main point of the fundamental theorem of linear algebra was to show the existence of Lagrange multipliers. The separating hyperplane theorem now fulfills that role instead. The fundamental theorem of linear algebra is somewhat easier to prove than the separating hyperplane theorem, but the fundamental theorem of linear algebra only applies to finite dimensional spaces, so perhaps it is not so "fundamental" afterall.

Definition B.1. If $S_1 + \cdots + S_k = V$ and $\forall \mathbf{v} \in V$, there are unique $\mathbf{x}_j \in S_j$ such that

$$\mathbf{v} = \mathbf{x}_1 + \cdots + \mathbf{x}_k$$

then V is the **direct sum** of $S_1, ..., S_k$, written

$$V = S_1 \oplus \cdots \oplus S_k$$

Representing a vector space as a direct sum will be important for some of the results below. The following lemma will be useful.

Lemma B.1. *Suppose S_1 and S_2 are linear subspaces of V. Then $V = S_1 \oplus S_2$ iff $V = S_1 + S_2$ and $S_1 \cap S_2 = \{0\}$.*

Proof. Suppose $V = S_1 \oplus S_2$. Then by definition $V = S_1 + S_2$. Also, if $\mathbf{x} \in S_1 \cap S_2$, then $0 = \mathbf{x} + (-\mathbf{x})$. The definition of direct sum requires this representation to be unique, so $\mathbf{x} = 0$ must be the only element of $S_1 \cap S_2$.

Suppose $V = S_1 + S_2$ and $S_1 \cap S_2 = \{0\}$. Let $\mathbf{v} \in V$. Since $V = S_1 + S_2$, $\exists \mathbf{x}_1 \in S_1$ and $\mathbf{x}_2 \in S_2$ such that $\mathbf{v} = \mathbf{x}_1 + \mathbf{x}_2$. Suppose that $\mathbf{y}_1 \in S_1$ and $\mathbf{y}_2 \in S_2$ and $\mathbf{v} = \mathbf{y}_1 + \mathbf{y}_2$. Subtracting,

$$0 = \underbrace{(\mathbf{x}_1 - \mathbf{y}_1)}_{\in S_1} + \underbrace{(\mathbf{x}_2 - \mathbf{y}_2)}_{\in S_2}$$

so $(\mathbf{x}_1 - \mathbf{y}_1) = -(\mathbf{x}_2 - \mathbf{y}_2)$. By the definition of subspaces, S_1 and S_2 are closed under scalar multiplication. Therefore, $(\mathbf{x}_i - \mathbf{y}_i) \in S_1 \cap S_2$ for $i = 1, 2$. By assumption, this intersection only contains 0, so we can conclude that $\mathbf{x}_i = \mathbf{y}_i$. The representation of $\mathbf{v} = \mathbf{x}_1 + \mathbf{x}_2$ is unique. □

The next three lemmas will be used later when proving theorem B.1. You may want to skip ahead and only come back to these lemmas when they're needed.

Lemma B.2. *Suppose V is finite dimensional and S is a subspace of V. Then $\exists$ another subspace, W, of V such that $V = S \oplus W$, and $\dim(V) = \dim(S) + \dim(W)$.*

Proof. Construct a basis for S as follows. Set the basis $B = \{\}$. If $S = \text{span}(B)$, then stop. Otherwise, choose $\mathbf{b}_j \in S \setminus \text{span}(B)$ and add it to B. Since V is finite dimensional, this process must stop after at most $\dim(V)$ steps. This gives a basis B for S.

We will now construct a basis for W. Set $E = \{\}$. If $\text{span}(B \cup E) = V$, then stop. Otherwise choose $\mathbf{e}_j \in V \setminus \text{span}(B \cup E)$ and add it to E. Again, this process must stop because V is finite dimensional. Let $W = \text{span}(E)$. By construction $\text{span}(B \cup E) = S + W = V$. Also, lemma 1.4 implies that each $\mathbf{v} \in V$ can be uniquely written as a linear combination of elements of $B \cup E$, each $\mathbf{s} \in S$ can be uniquely written as a linear combination of B, and each $\mathbf{w} \in W$ can be uniquely written as a linear combination of E. It follows that for each $\mathbf{v} \in V$ there are unique $\mathbf{s} \in S$ and $\mathbf{w} \in W$ such that $\mathbf{v} = \mathbf{s} + \mathbf{w}$. □

Lemma B.3. *Suppose V is finite dimensional and S and T are subspaces of V. If $S \cap T = 0$ and $\dim(S) + \dim(T) = \dim(V)$, then $V = S \oplus T$.*

Proof. Let $b_1, ..., b_n$ be a basis for S and $e_1, ..., e_m$ be a basis for T. Suppose

$$\sum_{i=1}^{n} \alpha_i b_i + \sum_{j=1}^{m} \beta_j e_j = 0$$

Then let

$$\mathbf{x} = \sum_{i=1}^{n} \alpha_i b_i = -\sum_{j=1}^{m} \beta_j e_j$$

$\mathbf{x} \in S$, and $\mathbf{x} \in T$. We assume $S \cap T = 0$, so then $\mathbf{x} = 0$. Since $b_1, ..., b_n$ are linearly independent, we must have $\alpha_1 = \cdots = \alpha_n = 0$. Similarly all $\beta_j = 0$. Therefore, $b_1, ..., b_n$,

$e_1, \ldots, e_m$ are linearly independent. This is a linearly independent set of $\dim(S) + \dim(T) = \dim(V)$ elements, so it is a basis for V. Hence, $V = S + T$. □

Lemma B.4. *Let S and T be subspaces of a finite dimensional space V. Then*

$$\dim(S + T) = \dim(S) + \dim(T) - \dim(S \cap T).$$

Proof. We will be brief for this proof. You may want to add the details as an exercise.

Let B be basis for $S \cap T$. Extend U to a basis for S, call it $B \cup B_S$. Similarly extend U to a basis for T, $B \cup B_T$. Then $\text{span}(B \cup B_S \cup B_T) = S + T$. Also, you can show that $B \cup B_S \cup B_T$ is linearly independent, and hence a basis, for $S + T$. Finally, note that since B, B_S, and B_T must be disjoint,

$$\dim(S + T) = |B \cup B_S \cup B_T| = |B \cup B_S| + |B \cup B_T| - |B|.$$

□

There is an interesting relationship among the null spaces and ranges of a linear transformation and its transpose. Let $A \in L(V, W)$ and suppose V and W are finite dimensional. Finite dimension ensures that $V^* = V$ and $W^* = W$. Then $\text{null}A \subseteq V$ and $\text{range}A^T \subseteq V$. How are these subspaces related?

Theorem B.1 (Fundamental theorem of linear algebra)**.** *Let $A \in L(V, W)$, where V and W are finite dimensional. Then*

$$V = \text{null}A \oplus \text{range}A^T$$

and

$$W = \text{null}A^T \oplus \text{range}A.$$

Also, $\dim(\text{range}A) = \dim(\text{range}A^T)$ *and,*

$$\dim(V) = \dim(\text{null}A) + \dim(\text{range}A)$$

and

$$\dim(W) = \dim(\text{null}A^T) + \dim(\text{range}A).$$

Proof. Suppose x is in the null space of A. Then $Ax = 0$. By the definition of the transpose,

$$w^T Ax = (A^T w)^T x = 0$$

The set $\{v : v = A^T w\}$ is the range of A^T. If $x \in \text{null}A \cap \text{range}A^T$, then there is a w such that $A^T w = x$. We know from the previous equation that $(A^T w)^T x = 0 = x^T x$. Therefore, it must be that $x = 0$, so $\text{null}A \cap \text{range}A^T = \{0\}$.

Let U be such that $\text{null}A \oplus U = V$ (lemma B.2), and $\dim(\text{null}A) + \dim(U) = \dim(V)$. Theorem 5.1 then implies that, $\dim(U) = \dim(\text{range}A)$. Additionally, lemma B.4 implies that $\dim(\text{null}A + \text{range}A^T) = \dim(\text{null}A) + \dim(\text{range}A^T) \leq \dim(V)$, so we can conclude that $\dim(\text{range}A^T) \leq \dim(\text{range}A)$.

Identical reasoning shows the opposite inequality. Therefore $\dim(\text{range}A) = \dim(\text{range}A^T)$. Finally, since $\dim(\text{null}A) + \dim(\text{range}A^T) = \dim(V)$ and $\text{null}A \cap \text{range}A^T = \{0\}$, using lemma B.3, we can conclude that $V = \text{null}A \oplus \text{range}A^T$. Identical reasoning shows that $W = \text{null}A^T \oplus \text{range}A$. □

Strang (1993) has a nice discussion of this theorem.

The requirement of finite dimension is really essential in this theorem. In infinite dimension, V^* need not be the same as V. If V and V^* are different, it does not make sense to talk about null$A$ + rangeA^T, because null$A \subseteq V$ and range$A^T \subseteq V^*$. Some people object to calling this theorem "fundamental" because it does not extend to infinite dimensional spaces.

This theorem has some nice implications for systems of linear equations. If $A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$ is a matrix, then the range of A is

$$\{Ax : x \in \mathbb{R}^n\} = \left\{ \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \vdots \sum_{j=1}^n a_{mj}x_j \end{pmatrix} : x_j \in \mathbb{R} \right\},$$

the set of linear combinations of the columns of A. The **column space** of A, denoted Col(A), is the space spanned by the column vectors of A. The column space of A is the same as the range of A.

Similar reasoning shows that the space of linear combinations of the rows of A is the range of A^T. The **row space** of A, denoted Row(A), is the space spanned by the row vectors of A. The fundamental theorem of linear algebra shows that the dimensions of the column and row spaces are equal. The rank of a matrix is the dimension of its row and column spaces.

Definition B.2. The **rank** of a linear transformation is the dimension of its range.

Example B.1. Let X be an $n \times k$ matrix. Define:

$$P_x = X(X^TX)^{-1}X^T$$

and

$$M_x = I - X(X^TX)^{-1}X^T.$$

Both P_x and M_x are linear transformations from $\mathbb{R}^n$ to $\mathbb{R}^n$. Also, both $P_x = P_x^T$ and $M_x = M_x^T$. Therefore, from the fundamental theorem of linear algebra,

$$\mathbb{R}^n = \text{null}P_x \oplus \text{range}P_x$$

and

$$\mathbb{R}^n = \text{range}M_x \oplus \text{null}M_x.$$

Suppose $w \in$ rangeP_x, then $\exists y \in \mathbb{R}^n$ such that $P_xy = w$. Notice that

$$\begin{aligned} M_xw =& M_xP_xy \\ =& (P_x - P_x)y = 0, \end{aligned}$$

so $w \in$ nullM_x. Similarly, if $w \in$ rangeM_x, then $w \in$ nullP_x. Also, if $w \in$ nullP_x, then

$$M_xw = w - P_xw = w$$

so $w \in \text{range} M_x$. We can conclude that null P_x = range M_x and null M_x = range P_x. Hence,

$$\begin{array}{ccccc} \mathbb{R}^n = & \text{null} P_x & \oplus & \text{range} P_x \\ & \parallel & & \parallel \\ \mathbb{R}^n = & \text{range} M_x & \oplus & \text{null} M_x. \end{array}$$

In studying constrained optimization problems,

$$\max_x f(x) \text{ s.t. } h(x) = c,$$

we used the fact that

$$Dh_x v = 0 \implies Df_x v = 0 \tag{5}$$

is equivalent to there existing λ^T such that

$$Df_x + \lambda^T Dh_x = 0.$$

Equation (5) is just another way of saying $\text{null} Dh_x \subseteq \text{null} Df_x$. An immediate consequence of theorem B.1 is the following.

Corollary B.1. *Let $A \in L(V, W)$ and $B \in L(V, Z)$. Then* $\text{null} A \subseteq \text{null} B$ *iff* $\text{range} A^T \supseteq \text{range} B^T$.

Proof. From theorem B.1, we know that $V = \text{null} A \oplus \text{range} A^T = \text{null} B \oplus \text{range} B^T$. Suppose $\text{null} A \subseteq \text{null} B$. Let $v \in \text{range} B^T$. Then either $v = 0$, in which case $v \in \text{range} A^T$, or $v \notin \text{null} B$. If $v \notin \text{null} B$, $v \notin \text{null} A$ as well. Therefore, $v \in \text{range} A^T$.

If $\text{range} A^T \supseteq \text{range} B^T$, an identical argument shows $\text{null} A \subseteq \text{null} B$. □

Letting $A = Dh_x$ and $B = Df_x$, we know that $\text{null} Dh_x \subseteq Df_x$ iff $\text{range} Dh_x^T \supseteq \text{range} Df_x^T$. This means that for every $w^* \in W^*$, $\exists \lambda_w \in Z^*$ such that

$$Df_x^T w^* = Dh_x^T \lambda_w.$$

For optimization problems, $W = \mathbb{R}$, so it suffices to just consider $w^* = 1$,

$$Df_x^T = Dh_x^T \lambda.$$

If V and Z are finite dimensional, then we can substract and transpose to get

$$Df_x - \lambda^T Dh_x = 0.$$

Corollary (B.1) is also true in infinite dimensional spaces under some additional conditions.[7] In fact, theorem 7.2 is equivalent to this corollary. When working with optimal control problems, we were using this result and making some implicit assumptions about how elements from the duals spaces can be written as integrals. Appropriately defining V and Z ensures that these assumptions hold. The details are tedious, so we will not go into them.

[7] The ranges of A and B must be closed. See e.g. section 6.6 of Luenberger (1969).

References

Axler, Sheldon. 1997. *Linear algebra done right*. Springer-Verlag.

Carter, Michael. 2001. *Foundations of mathematical economics*. MIT Press.

Clarke, Francis. 2013. *Functional analysis, calculus of variations and optimal control*, vol. 264. Springer.

De la Fuente, Angel. 2000. *Mathematical methods and models for economists*. Cambridge University Press.

Holmes, Richard B. 1975. *Geometric functional analysis and its applications*. Springer Science & Business Media.

Luenberger, David G. 1969. *Optimization by vector space methods*. John Wiley & Sons.

Simon, Carl P and Lawrence Blume. 1994. *Mathematics for economists*. Norton New York.

Strang, Gilbert. 1993. "The Fundamental Theorem of Linear Algebra." *The American Mathematical Monthly* 100 (9):pp. 848–855. URL http://www.jstor.org/stable/2324660.