MATRIX ALGEBRA AND INTRODUCTION TO VECTOR SPACES

WRITTEN BY PAUL SCHRIMPF AND MODIFIED BY HIRO KASAHARA

SEPTEMBER 2, 2020

UNIVERSITY OF BRITISH COLUMBIA

ECONOMICS 526

We have already used matrices in our study of systems of linear equations. Matrices are useful through economics. You will see a lot of matrices in econometrics. Understanding matrices is also essential for understanding systems of nonlinear equations, which appear in all fields of economics. This lecture will go over some fundamental properties of matrices. It is based on a combination of Simon and Blume Chapters 8-9, Appendix A of *Econometric Analysis* by William Green, and Chapter 9 of *Principles of Mathematical Analysis* by Rudin. Carter sections 1.4.1 and 1.4.2 covers much the same material as section 1.2 of these notes. The start of chapter 3 of Carter (up to 3.1.1) covers much the same material as section 1.2 and the start of section 2 of these notes.

Matrices are often introduced as just arrays of numbers. That sort of definition works out alright. However, it makes certain subsequent definitions somewhat mysterious. You might wonder why matrix multiplication is defined the way it is, or why transposing is important. Well, there is another, more abstract, but also more fundamental way of describing matrices. This way of looking at matrices will reveal exactly why matrix multiplication is the way it is, and illuminate many other properties of matrices.

1. VECTOR SPACES

To fully understand matrices, we must first under vector spaces. Loosely, 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 $(1.1y_1^A, 1.1y_2^A)$.

You are likely already familiar with $\mathbb{R}^n$, especially $\mathbb{R}^2$ and $\mathbb{R}^3$, which are the most common vector spaces. There are three ways of approaching vector spaces. The first is geometrically — introduce vectors as directed arrows 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.

Definition 1.1. A **vector space** is a set V and a field $\mathbb{F}$ 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 $\mathbb{F}$ 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 $f \in \mathbb{F}$ we have $vf \in V$
   (b) Distributivity: $\forall v_1, v_2 \in V$ and $f_1, f_2 \in \mathbb{F}$
$$f_1(v_1 + v_2) = f_1 v_1 + f_1 v_2$$
   and
$$(f_1 + f_2)v_1 = f_1 v_1 + f_2 v_1$$
   (c) Consistent with field multiplication: $\forall v \in V$ and $f_1, f_2 \in V$ we have
$$1v = v$$
   and
$$(f_1 f_2)v = f_1(f_2 v)$$

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

Example 1.1. $\mathbb{R}^n$ with the field $\mathbb{R}$ 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**[1] 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 $(V, \mathbb{F}, +, \cdot)$ when referring to a vector space.

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]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. Any linear subspace of $\mathbb{R}^n$ along with the field $\mathbb{R}$ and the usual vector addition and scalar multiplication is a vector space. Linear subspaces are closed under $+$ and $\cdot$ by definition. Linear subspaces inherit all the other properties required by the definition of a vector space from $\mathbb{R}^n$.

Example 1.3. $(\mathbb{Q}^n, \mathbb{Q}, +, \cdot)$ is a vector space where $+$ and $\cdot$ defined as in 1.1.

Example 1.4. $(\mathbb{C}^n, \mathbb{C}, +, \cdot)$ where $+$ and $\cdot$ defined as in 1.1 except with complex addition and multiplication taking the place of real addition and multiplication.

Except for the preceding two examples, all the vector spaces in this class will be real vector spaces with the field $\mathbb{F} = \mathbb{R}$. Two more examples of vector spaces that we have already encountered include:

Example 1.5. The set of all solutions to a system of linear equation with $b = 0$, i.e., $(x_1, ..., x_n) \in \mathbb{R}^n$ such that

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

and

Example 1.6. Any linear subspace of a vector space.

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

Example 1.7. 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.

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 1.8. 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 1.9. The set of all continuous functions with addition and scalar multiplication defined as in 1.8.

Example 1.10. The set of all k times continuously differentiable functions with addition and scalar multiplication defined as in 1.8.

Example 1.11. The set of all polynomials with addition and scalar multiplication defined as in 1.8.

Example 1.12. The set of all polynomials of degree at most d with addition and scalar multiplication defined as in 1.8.

Example 1.13. The set of all functions from $\mathbb{R} \to \mathbb{R}$ such that $f(29481763) = 0$ with addition and scalar multiplication defined as in 1.8.

Example 1.14. Let $1 \leq p < \infty$ and let $\mathcal{L}^p(0,1)$ be the set of functions from $(0,1)$ to $\mathbb{R}$ such that $\int_0^1 |f(x)|^p dx$ is finite. [2] Then $\mathcal{L}^p(0,1)$ with the field $\mathbb{R}$ and addition and scalar multiplication defined as

$$\begin{aligned}(f+g)(x) =& f(x) + g(x) \\ (\alpha f)(x) =& \alpha f(x)\end{aligned}$$

is a vector space. The only difficult part of the definition of vector spaces to verify is closure under addition. To prove closure under addition, we can use Jensen's inequality. This is a surprisingly useful inequality that we may or may not prove later. Anyway the simplest form of Jensen's inequality says that if $h(x)$ is convex[3], then for any $a_1 > 0$ and $a_2 > 0$,

$$h\left(\frac{a_1 x_1 + a_2 x_2}{a_1 + a_2}\right) \leq \frac{a_1 h(x_1) + a_2 h(x_2)}{a_1 + a_2}.$$

If $p > 1$, $h(x) = |x|^p$ is convex. From Jensen's inequality,

$$\begin{aligned} |\frac{f(x)+g(x)}{2}|^p \leq & \frac{|f(x)|^p + |g(x)|^p}{2} \\ |f(x)+g(x)|^p \leq 2^{p-1}|f(x)|^p + |g(x)|^p. & \end{aligned}$$

Integrating,

$$\int |f(x)+g(x)|^p dx \leq 2^{p-1}\left(\int |f(x)|^p dx + \int |g(x)|^p dx\right)$$

The right side is finite for any $f, g \in \mathcal{L}^p(0,1)$, so the left side is also finite and $f + g \in \mathcal{L}^p(0,1)$.

To a get a better feel for $\mathcal{L}^p(0,1)$, you should try to come with some examples of familiar functions are or are not included. You could consider polynomials, rational functions (ratios of polynomials), exponential, and logarithm.

[2] For now, you can just think of $\mathcal{L}^p(0,1)$ as consisting of all functions from $(0,1) \to \mathbb{R}$. The p and finite integral part only become important when we think of $\mathcal{L}^p$ as normed vector spaces, which we will do in the next set of notes.

[3] A real valued function $h(x)$ is (weakly) convex if for all x_1, x_2 and $\lambda \in (0,1)$ $h(\lambda x_1 + (1-\lambda)x_2) \leq \lambda h(x_1) + (1-\lambda)h(x_2)$.

1.1. Linear combinations.

Definition 1.2. Let V be a vector space and $v_1, ..., v_k \in V$. A **linear combination** of $v_1, ..., v_k$ is any vector

$$c_1 v_1 + ... + c_k v_k$$

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

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 $c_1 v_1 + ... + c_k v_k \in V$.

If we take all possible linear combinations of, $\{c_1 x_1 + c_2 x_2 : c_1 \in \mathbb{R}, c_2 \in R\}$, then the set will contain 0, and it will be a linear subspace. This motivates the following definition.

Definition 1.3. 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 $W = \{v_1, ..., v_k\}$, the span of W is the set

$$\{c_1 v_1 + ... + c_k v_k : c_1, ..., c_k \in \mathbb{F}\}.$$

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 1.8. The span of $\{1, x, ..., x^n\}$ is the set of all polynomials of degree less than or equal n.

You might remember the next two definitions from the previous lecture.

Definition 1.4. A set of vectors $W \in V$, is **linearly independent** if the only solution to

$$\sum_{j=1}^{k} c_j v_j = 0$$

is $c_1 = c_2 = ... = c_k = 0$ for any k and $v_1, ..., v_k \in W$.

Checking linear independence. Given a set of vectors, $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$ (or any n-dimensional vector space), how do check whether they are linearly independent? Well, by definition, they are linearly independent if $c_1 = c_2 = ... = c_n = 0$ is the only solution to

$$\sum_{j=1}^{n} c_j \mathbf{v}_j = 0$$

If we write this condition as a system of linear equations we have

$$\begin{aligned} v_{11} c_1 + v_{12} c_2 + ... + v_{1n} c_n &= 0 \\ \vdots &= \vdots \\ v_{m1} c_1 + v_{m2} c_2 + ... + v_{mn} c_n &= 0 \end{aligned}$$

or in matrix form,

$$\begin{pmatrix} v_{11} & \cdots & v_{1n} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{mn} \end{pmatrix} \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} = 0$$

$$V\mathbf{c} = 0$$

We call any system with 0 on the right hand side a **homogeneous** system. Any homogeneous system always has $\mathbf{c} = 0$ as a solution. We know from lecture 2 that it will have other solutions if the rank of V is less than n. This proves the following lemma.

Lemma 1.2. *Vectors $\mathbf{v}_1, ..., \mathbf{v}_n \in \mathbb{R}^m$ are linearly independent if and only if*

$$\text{rank}(\mathbf{v}_1, ..., \mathbf{v}_n) = n.$$

Corollary 1.1. *Any set of $k > m$ vectors in a $\mathbb{R}^m$ space are linearly dependent.*

1.1.1. *Dimension and basis.* Recall the definition of dimension from last class.

Definition 1.5. The **dimension** of a vector space, V, is the cardinality of the largest set of linearly independent elements in V.

Definition 1.6. 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, we will soon see that dimension of V must be exactly k.

Example 1.16. 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?

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

Lemma 1.3. *Let B be a basis for a vector space V. Then $\forall v \in V$ there exists a unique $v_1, ..., v_k \in \mathbb{F}$ and $b_1, ..., b_k \in B$ such that $v = \sum_{i=1}^k v_i b_i$*

Proof. By the definition of a basis, B spans V, so such $(v_1, ..., v_k)$ must exist. Now suppose there exists another such $(v'_1, ..., v'_j)$ and associated b'_i. The $\{b_1, ..., b_k\}$ and $\{b'_1, ..., b'_j\}$ might not be the same collection of elements of B. Let $\{\tilde{b}_1, ..., \tilde{b}_n\} = \{b_1, ..., b_k\} U \{b'_1, ..., b'_j\}$. Define $\tilde{v}_i = v_j$ if $\tilde{b}_i = b_j$, else 0. Similarly define $\tilde{v}'_i$. With this new notation we have

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

$$\sum (\tilde{v}_i - \tilde{v}'_i) \tilde{b}_i = 0$$

However, if B is a basis, its elements must be linearly independent so $\tilde{v}_i = \tilde{v}'_i$ for all i, so the original $v_1, ..., v_k$ must be unique. $\square$

Let us show that dimension and basis are sensible definitions by showing that any two bases have the same size.

Lemma 1.4. *If B is a basis for a vector space V and $I \subseteq V$ is a set of linearly independent elements then $|I| \leq |B|$.*

For vector spaces of finite dimension this lemma can be stated as follows. If $b_1, ..., b_m$ is a basis for a vector space V and $v_1, ..., v_n \in V$ are linearly independent, then $n \leq m$. We will only prove this finite dimensional version, but it is true for infinite dimensional spaces as well.

Proof. Write $v_i \in I$ in terms of the basis:

$$v_i = \sum_{i=1}^{m} a_{ij} b_i$$

Let us check for linear independence as described above.

$$0 = \underbrace{A}_{m \times n} \underbrace{c}_{n \times 1}$$

This is a system of m equations with n unknowns. If $n > m$, then it must have multiple solutions, and $v_1, ..., v_n$ cannot be linearly independent. □

An immediate consequence of this lemma is that any two bases must have the same cardinality.

Corollary 1.2. *Any two bases for a vector space have the same cardinality.*

Proof. If not, one would have more elements than the other. That would contradict the previous lemma. □

Example 1.17. What is the dimension of each of the examples of vector spaces above? Can you find a basis for them?

We have now shown that any basis of a vector space has cardinality equal to the dimension of the space. A related useful fact is that if we have n elements that span an n dimensional space, then those elements must be linearly independent, and hence a basis.

Lemma 1.5. *If $v_1, ..., v_n$ span an n dimensional vector space, then $v_1, ..., v_n$ are linearly independent.*

Proof. If $v_1, ..., v_n$ span $\mathbb{R}^n$, then we can write each standard basis vector as

$$e_i = \sum_{j=1}^{n} a_{ij} v_j.$$

Suppose we have $c_j \neq 0$ such that

$$0 = \sum_{j=1}^{n} c_j v_j$$

$$v_n = \sum_{j=1}^{n-1} -\frac{c_j}{c_n} v_j.$$

Substituting,

$$e_i = \sum_{j=1}^{n-1} (a_{ij} - \frac{c_j}{c_n} a_{in}) v_j.$$

Writing these n systems of linear equations in matrix form we have

$$I_n = \underbrace{A}_{n\times(n-1)} \underbrace{V}_{(n-1)\times n}.$$

However, we know that rank$I_n = n$ and rank$(AV) \leq \max\{\text{rank}A, \text{rank}V\} \leq n-1$, so this equation cannot be true. The steps to get to this equation only relied on some $c_j \neq 0$ so that we could divide by it. Therefore, it must be that $c_j = 0$ for all j, and $v_1, ..., v_n$ are linearly independent. □

Collecting the results above and specializing them to $\mathbb{R}^n$ we have four equivalent ways of describing a basis.

Theorem 1.1. *Let $v_1, ..., v_n \in \mathbb{R}^n$ and let $V = (v_1, ..., v_n)$ be the n by n matrix with columns v_i. Then the following are equivalent:*

(1) *$v_1, ..., v_n$ are linearly independent,*
(2) *$v_1, ..., v_n$ span $\mathbb{R}^n$,*
(3) *$v_1, ..., v_n$ are a basis for $\mathbb{R}^n$,*
(4) rank$V = n$

Suppose V is an n-dimension real[4] 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 $v \in V$, there are unique $v_1, ..., v_n \in \mathbb{R}$ such that

$$v = \sum_{i=1}^{n} v_i b_i.$$

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

$$I(v) = (v_1, ..., v_n).$$

By lemma 1.3, I must be one-to-one. I must also be onto since by definition of a vector space, for any $(v_1, ..., v_n) \in \mathbb{R}^n$, the linear combination, $\sum_{i=1}^n v_i b_i$ is in V. Moreover, I preserves addition in that for any $v^1, v^2 \in V$,

$$\begin{aligned} I(v^1 + v^2) =& (v_1^1 + v_1^2, ..., v_n^1 + v_n^2) \\ =& (v_1^1, ..., v_n^1) + (v_1^2 + ... + v_n^2) \\ =& I(v^1) + I(v^2). \end{aligned}$$

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

$$I(\alpha v) = \alpha I(v).$$

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.

[4]Meaning the field of scalars is $\mathbb{R}$.

Definition 1.7. 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, $I : V \to W$ such that

$$I(v^1 + v^2) = I(v^1) + I(v^2)$$

for all $v^1, v^2 \in V$, and

$$I(\alpha v) = \alpha I(v)$$

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

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

Theorem 1.2. *Let V be an n-dimensional vector space over the field $\mathbb{F}$. Then V is isomorphic to $\mathbb{F}^n$.*

Proof. We proved this for $\mathbb{F} = \mathbb{R}$. The same argument works for any field. □

2. 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 2.1. A **linear transformation** (aka linear function) is a function, A, from a vector space $(V, \mathbb{F}, +, \cdot)$ to a vector space $(W, \mathbb{F}, +, \cdot)$ such that $\forall v_1, v_2 \in V$,

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

and

$$A(fv_1) = fAv_1$$

for all scalars $f \in \mathbb{F}$.

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 2.1. 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 when we were working with systems of linear equations, 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 $A(f\mathbf{x}_1 + \mathbf{x}_2) = fA\mathbf{x}_1 + \mathbf{x}_2$ for scalars $f \in \mathbb{R}$ 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 $v_j \in \mathbb{F}$. 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 $d_1, ..., d_m$ 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 $d_1, ..., d_m$, 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 $a_{ij} \in \mathbb{F}$ 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 $a_{ij} \in \mathbb{F}$ we can construct a linear transformation. This leads us to the following result.

Theorem 2.1. *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 2.2 (Integral operator)**.** Let $k(x,y)$ be a function from $(0,1)$ to $(0,1)$ such that $\int_0^1 \int_0^1 k(x,y)^2 dxdy$ 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 2.3 (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

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

and

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

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 2.4 (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 will encounter some linear transformations on infinite dimensional spaces when we study optimal control. However, for the rest of this lecture, our main focus will be an linear transformations between finite dimensional spaces, i.e. matrices.

3. 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 $Be_j = \sum_{i=1}^m b_{ij}e_i$.

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

3.2. **Scalar multiplication.** The definition of linear transformations requires that $A\alpha x = \alpha Ax$ where $\alpha \in \mathbb{F}$ 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 3.1. 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 3.2. 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.

3.3. **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 transformation acts on a basis vectors, so to define multiplication, we should look at $A(Be_k)$

$$\begin{aligned}
A(Be_k) =& A(\sum_{j=1}^n b_{jk}e_j) && \text{definition of } Be_k \\
=& \sum_{j=1}^n b_{jk}Ae_j && \text{Definition of linear transformtion} \\
=& \sum_{j=1}^n b_{jk}\left(\sum_{i=1}^m a_{ij}e_i\right) && \text{definition of } Ae_j \\
=& \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij}b_{jk}\right) e_i \\
=& \begin{pmatrix} \sum_{j=1}^n a_{1j}b_{j1} & \cdots & \sum_{j=1}^n a_{1j}b_{jp} \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^n a_{mj}b_{j1} & \cdots & \sum_{j=1}^n a_{mj}b_{jp} \end{pmatrix} e_k \\
=& (AB)e_k.
\end{aligned}$$

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 $I_n x = x \forall x \in \mathbb{R}^n$.

As we saw on the review, matrix multiplication is not commutative.

3.4. **Transpose.** As you likely know, the transpose of an $m \times n$ matrix A is an $n \times m$, A^T whose rows are the columns of A. As a linear transformation, if $A : V \to W$, then $A^T : W \to V$, but how can we define A^T when A cannot be represented as a matrix (i.e. when V or W is infinite dimensional). There are two ways to define the transpose of a linear function. They each require introducing a new concept. We will first define the transpose in terms of the inner product. We will then define the transpose in terms of dual spaces. Both dual spaces and inner products will appear later in the course as well, so its worth introducing them now. However, it is not essential to completely understand either of these abstract definitions of the transpose. You will only really need to understand the transposing of matrices.

3.4.1. *Transpose and inner products.* One way to define the transpose in terms of linear transformations, is to use another property of Euclidean space ($\mathbb{R}^n$).

Definition 3.1. A real **inner product space** is a vector space over the field $\mathbb{R}$ with an additional operation called the inner product that is function from $V \times V$ to $\mathbb{R}$. We denote the inner product of $v_1, v_2 \in V$ by $\langle v_1, v_2 \rangle$. It has the following properties:

(1) Symmetry: $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$
(2) (Bi)linear: $\langle av_1 + bv_2, v_3 \rangle = a \langle v_1, v_3 \rangle + b \langle v_2, v_3 \rangle$ for $a, b \in \mathbb{R}$
(3) Positive definite: $\langle v, v \rangle \geq 0$ and equals 0 iff $v = 0$.

Although it is not obvious from this definition, an inner product space is a vector space where we can measure the angle between two vectors. We will come back to this point next lecture.

Example 3.3. $\mathbb{R}^n$ with the **dot product**, $x \cdot y = \sum_{i=1}^n x_i y_i$, is an inner product space.

Example 3.4. As it is usually defined, $\mathcal{L}^2(0,1)$ with $\langle f, g \rangle \equiv \int_0^1 f(x)g(x)dx$ is an inner product space. However, we have not defined $\mathcal{L}^2(0,1)$ carefully enough for this to be an inner product space. For example, consider $h(x) = \begin{cases} 0 \text{ if } x \neq 1/2 \\ 1 \text{ if } x = 1/2 \end{cases}$. Then

$$\langle h, h \rangle = \int_0^1 h(x)^2 dx = 0,$$

but $h(x)$ is not zero, violating the positive definite requirement of inner products. We can get around this problem by defining $\mathcal{L}^2(0,1)$ as not all functions, but as equivalence classes of functions. We say $f \simeq g$ if $\int_0^1 |f(x) - g(x)|^2 dx = 0$, and then define $\mathcal{L}^2(0,1)$ as the set of equivalence classes of functions. With this definition, the problematic h above and 0 represent the same element of $\mathcal{L}^2(0,1)$, and it becomes an inner product space.

As an exercise, you could try to verify some of these assertions. For example, show that $\simeq$ is really an equivalence relation, or show that $\mathcal{L}^2(0,1)$ remains a vector space when it is the set of equivalence classes instead of set of functions. Fully verifying these claims requires knowledge of measure theory, which is outside the scope of this course, but you could make some progress.

We can now define the transpose in terms of the inner product.

Definition 3.2. Given a linear transformation, A, from a real inner product space V to a real inner product space W, the **transpose** of A, denoted A^T (or often A') is a linear transformation from W to V such that $\forall v \in V, w \in W$

$$\langle Av, w \rangle = \left\langle v, A^T w \right\rangle.$$

Now, let's take $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$ and see what this definition means for the transpose of a matrix. First observe that A^T is a mapping from $\mathbb{R}^m$ to $\mathbb{R}^n$, so as a matrix is must be n by m. Let a^T_{ji} denote the entries of A^T. Rewriting the inner product in terms of the entries of the matrix, we get

$$\begin{aligned}\langle Av, w \rangle &= \sum_{i=1}^m \left(\sum_{j=1}^n a_{ij} v_j \right) w_i \\ &= \sum_{i=1}^m \sum_{j=1}^n a_{ij} w_i v_j\end{aligned}$$

and

$$\begin{aligned}\left\langle v, A^T w \right\rangle &= \sum_{j=1}^n v_j \left(\sum_{i=1}^m a^T_{ji} w_i \right) \\ &= \sum_{i=1}^m \sum_{j=1}^n a^T_{ji} w_i v_j\end{aligned}$$

If $\langle Av, w \rangle = \langle v, A^T w \rangle$, for any v and w we must have $a^T_{ji} = a_{ij}$.

The transpose of a matrix simply swaps rows for columns. The transpose has the following properties:

(1) $(A+B)^T = A^T + B^T$
(2) $(A^T)^T = A$
(3) $(\alpha A)^T = \alpha A^T$
(4) $(AB)^T = B^T A^T$.
(5) $\text{rank} A = \text{rank} A^T$

3.4.2. *Transpose and dual spaces.* Another way of defining the transpose of a linear transformation is in terms of dual spaces. Not all vector spaces are inner product spaces, but all vector spaces have a dual space, so defining the transpose using dual spaces will be more general. Dual spaces are needed to define differentiation on vector spaces, which we will want to do later.

Definition 3.3. Let V be a vector space. The **dual space** of V, denote V^* is the set of all (continuous)[5] linear functionals, $v^* : V \to \mathbb{R}$.

Example 3.5. 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.

Example 3.6. 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.

Example 3.7. What is the dual space of $V = \mathcal{L}^2(\mathbb{R}, f_x) = \{g : \mathbb{R} \to \mathbb{R}$ 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 $g \in V$. H older's inequality[6] , which we have not studied but is good to be aware of, says that

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

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^*.[7]

We will have to work with V^* and similar dual spaces when we study optimal control.

[5] We have not yet defined continuity, so do not worry about this requirement. All linear functionals on finite dimensional spaces are continuous. Some linear functionals on infinite dimensional spaces are not continuous. The definition of dual space does not always require continuity. Often 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. We will ignore this distinction.

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

[7] This is a consequence of the Radon-Nikodym theorem, which is beyond the scope of this course.

Definition 3.4. If $A : V \to W$ is a linear transformation, then the **transpose** (or dual) 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 $w^*(Av) \in \mathbb{R}$.

Let us verify that this definition of the transpose coincides with the one using inner products.[8] Let V and W be innerproduct spaces and $A : V \to W$ be linear, and suppose $A^T : W \to V$ satisfies the earlier definition of transpose, that is $\langle Av, w\rangle = \langle v, A^T w\rangle$. We want to show that A^T also satisfies the later definition transpose. First, we must establish that for inner product spaces $W^* = W$ and $V^* = V$. For each $v \in V$ we can define a linear operator $v^*(y) = \langle v, y\rangle$. Thus, $V \subseteq V^*$. When $V = \mathbb{R}^n$, it is clear that the only linear operators on V are of the form $a_1 x_1 + ... + a_n x_n$, so in that case $V^* = V$. In an example we argued that $\mathcal{L}^{2^*} = \mathcal{L}^2$. We will take as given for the remainder of the course $V^* = V$ whenever V is an inner product space[9]. Given this, we can conclude that

$$w^*(Av) = \langle Av, w\rangle = \left\langle v, A^T w \right\rangle = (A^T w^*)(v).$$

As an exercise, you should try to show that this definition of the transpose is the same as the familiar flipping rows and columns definition for matrices.

3.5. **Types of matrices.** There are some special types of matrices that will be useful.

Definition 3.5. A **column** matrix is any m by 1 matrix.

Definition 3.6. A **row** matrix is any 1 by n matrix.

Definition 3.7. A **square** matrix has the same number of rows and columns.

Definition 3.8. A **diagonal** matrix is a square matrix with non-zero entries only along its diagonal, i.e. $a_{ij} = 0$ for all $i \neq j$.

Definition 3.9. An **upper triangular** matrix is a square matrix that has non-zero entries only on or above its diagonal, i.e. $a_{ij} = 0$ for all $j > i$. A **lower triangular** matrix is the transpose of an upper triangular matrix.

Definition 3.10. A matrix is **symmetric** if $A = A^T$.

Definition 3.11. A matrix is **idempotent** if $AA = A$.

Definition 3.12. A **permutation** matrix is a square matrix of 1's and 0's with exactly one 1 in each row or column.

Definition 3.13. A **nonsingular** matrix is a square matrix whose rank equals its number of columns.

[8] The two definitions coincide for $V = \mathbb{R}^n$ and $W = \mathbb{R}^m$. There are some situations when the definitions are not the same, but we will not encounter any. When you care about the difference the previous definition of transpose using inner products is usually called the (Hermitian) adjoint, and the definition using dual spaces is called the transpose.

[9] The requirement of continuity in our definition of dual space is essential for this to be true of infinite dimensional V.

Definition 3.14. An **orthogonal** matrix is a square matrix such that $A^T A = I$.

3.6. **Invertibility.** We will now define the inverse of a matrix and linear transformation. Since multiplication of linear transformations is not commutative, it could be that right and left inverses differ. To allow this possibility, we make the following definition.

Definition 3.15. Let A be a linear transformation from V to W. Let B be a linear transfromation from W to V. B is a **right inverse** of A if $AB = I_V$. Let C be a linear transformation from V to W. C is a **left inverse** of A if $CA = I_W$.

We will now show that right and left inverses do not differ for square matrices.

Lemma 3.1. *If A is a linear transformation from V to V and B is a right inverse, and C a left inverse, then $B = C$.*

Proof. Left multiply $AB = I$ by C to get

$$\begin{aligned} CAB &= CI \\ (CA)B &= C \\ B &= C. \end{aligned}$$

□

Now that we have established that right and left inverses are the same, it makes sense to define just the inverse.

Definition 3.16. Let A be a linear transformation from V to V. The **inverse** of A is written A^{-1} and satisfies $AA^{-1} = I = A^{-1}A$.

The previous lemma shows that if A^{-1} exists, it is unique.

Lemma 3.2. *Let A be a linear tranformation from V to V, and suppose A is invertible. Then A is nonsingular and the unique solution to $Ax = b$ is $x = A^{-1}b$.*

Recall that A is nonsingular if $Ax = b$ has a unique solution for all b.

Proof. Multiply both sides of the equation by A^{-1} to get

$$\begin{aligned} A^{-1}Ax &= A^{-1}b \\ x &= A^{-1}b \end{aligned}$$

Therefore a solution exists. This solution is also unique because any other solution, x' must also satisfy

$$\begin{aligned} Ax' &= b \\ A^{-1}Ax' &= A^{-1}b \\ x' &= A^{-1}b, \end{aligned}$$

and A^{-1} is unique. □

Conversely we can show that if A is nonsingular, then A must be invertible.

Lemma 3.3. *If A is nonsingular, then A^{-1} exists.*

Proof. Let $e_1, e_2, ...$ be a basis for V. By the definition of nonsingular for any e_j there exists a solution to

$$Ax = e_j.$$

Call this solution x_j. Define a linear transformation C by

$$Cv = \sum v_j x_j,$$

where $v = \sum v_j e_j$. We can write any $v \in V$ in this way since e_j is a basis. Now, consider $A(Cv)$.

$$\begin{aligned} A(Cv) =& A \sum v_j x_j \\ =& \sum v_j A x_j \\ =& \sum v_j e_j = v. \end{aligned}$$

Thus, $AC = I$ and $C = A^{-1}$. □

Corollary 3.1. *A square matrix A is invertible if and only if* rank$A$ *is equal to its number of columns.*

Proof. In the previous lecture, we showed that rank$A =$ number of columns if and only if A is nonsingular. That result along with the two previous lemmas implies the corollary. □

Let A and B be invertible square matrices. The inverse has the following properties:

(1) $(AB)^{-1} = B^{-1}A^{-1}$
(2) $(A^T)^{-1} = (A^{-1})^T$
(3) $(A^{-1})^{-1} = A$

4. Determinants

I am not a fan of determinants because they do not generalize well to infinite dimensions. Nonetheless, you must at least know that the determinant of a matrix is non-zero if and only if the matrix is invertible.

Given a 1 by 1 matrix, say, $A = (a)$, we known that A^{-1} exists if and only if $a \neq 0$. Also, viewed as a linear transformation from $\mathbb{R}$ to $\mathbb{R}$, A transforms 1 into a. It would be useful to have a number with similar properties for larger matrices. That is, for any matrix A, we want a number that is not zero if and only if A^{-1} exists, and that describes something about how A acts as a linear transformation. Well, the determinant of a matrix, written det$A$ is exactly such a number.

Let us start by looking at a 2 by 2 matrix, $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. The inverse of A^{-1} must satisfy

$$AA^{-1} = I \tag{1}$$

If x_1 is the first column of A^{-1} and x_2 is the second column, then we can rewrite this as two systems of linear equations,

$$Ax_1 = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

and

$$Ax_2 = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

If you think about how we would solve these two systems using Gauss-Jordan elimination, you will see that the steps involved depend only on A and not on the right side of the equation. Therefore, we can solve both systems together by performing Gauss-Jordan elmination on

$$\begin{aligned} \begin{pmatrix} a & b & 1 & 0 \\ c & d & 0 & 1 \end{pmatrix} &\simeq \begin{pmatrix} a & b & 1 & 0 \\ 0 & \frac{ad-bc}{a} & -\frac{c}{a} & 1 \end{pmatrix} \\ &\simeq \begin{pmatrix} a & b & 1 & 0 \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} \\ &\simeq \begin{pmatrix} a & 0 & \frac{ad}{ad-bc} & \frac{-ba}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix} \\ &\simeq \begin{pmatrix} 1 & 0 & \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ 0 & 1 & -\frac{c}{ad-bc} & \frac{a}{ad-bc}. \end{pmatrix} \end{aligned}$$

So, $A^{-1} = \begin{pmatrix} \frac{d}{ad-bc} & \frac{-b}{ad-bc} \\ -\frac{c}{ad-bc} & \frac{a}{ad-bc} \end{pmatrix}$. For this to exist, we must have $ad - bc \neq 0$. (The steps to get the formula for A^{-1} divided by a, so we also assumed $a \neq 0$. However, if $a = 0$, we could start Gaussian elimination by swapping the first and second rows, divide by c, and end up with the same formula. If $c = 0$ as well, then $ad - bc = 0$ too, so $ad - bc \neq 0$ is really the only condition needed).

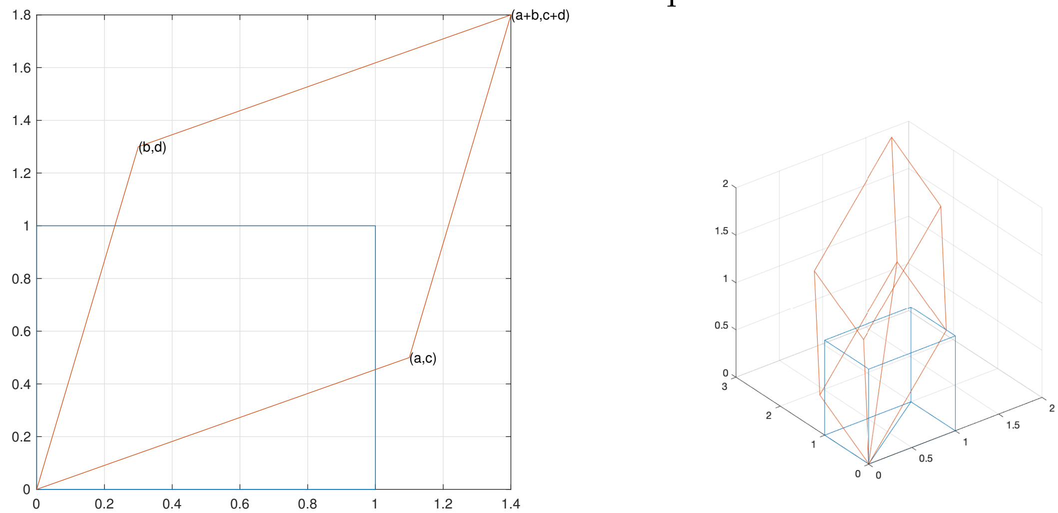
This calculation suggests that $ad - bc$ is a good candidate for the determinant of 2 by 2 matrices. You may remember from the review problem set that $|ad - bc|$ is the area of the parallelogram constructed from the columns of A. A nice interpretation of this parallelogram is that it is the image of the unit square under the linear transformation represented by A. Additionally, if $\det A > 0$, the transformation only stretches and shrinks the unit squares. If $\det A < 0$, the transformation also reflects the unit square over the line $x = -y$. Similarly, in higher dimensions you can consider the image of the unit cube (in 3d) or unit hypercube (i.e. higher dimensional cube). This image will be a higher-dimension analog of a parallelogram, and you could calculate its volume. It will turn out that $|\det A|$ is the volume of the hyper-parallelogram.[10]

Definition 4.1. Let A be an n by n matrix consisting of column vectors $a_1, ..., a_n$. The determinant of A is the unique function such that

(1) $\det I_n = 1$.

[10]Not a standard name.

FIGURE 1. Transformed unit square and unit cube



(2) As a function of the columnes, det is an alternating form: $\det(A) = 0$ iff $a_1, ..., a_n$ are linearly dependent.
(3) As a function of the columnes, det is multi-linear:

$$\det(a_1, ..., ba_j + cv, ..., a_n) = b\det(A) + c\det(a_1, ..., v, ...a_n)$$

Condition (1) seems very natural. Also, if we want $\det A$ to be the volume of the transformed unit cube, we must have $\det I_n = 1$. (2) ensures that $\det A = 0$ whenever A is not invertible.

Lemma 4.1. *Let A be an n by n matrix. The A is singular if and only if the columns of A are linearly dependent.*

Proof. If A is singular, then $\text{rank} A < n$. Also $\text{rank} A^T = \text{rank} A$. Therefore there must exist row operations such that the last row of A^T is all 0's. Row operations are interchanging rows, rescaling by a scalar, and adding multiples of one row to another. Therefore, there must be $c_1, ..., c_n \in \mathbb{R}$ such that

$$c_1 a_1 + ... + c_n a_n = 0$$

where a_i are the rows of A^T, which are also the columns of A.

Similarly if the columns of A are linearly dependent, then there are $c_1, ..., c_n \in \mathbb{R}$ such that

$$c_1 a_1 + ... + c_n a_n = 0.$$

These are all row operations, so we can perform them to make the last row of A^T all 0's. We can then perform Gaussian elimination on the first $n - 1$ rows of A^T to put it into row echelon form with the last row all 0's. Therefore, $\text{rank} A < n$ and A is singular. $\square$

Corollary 4.1. *A is nonsingular if and only if* $\det A \neq 0$.

The third condition on the determinant, 3, guarantee that $\det A$ can be interpreted as the volume of the transformed unit cube. It is easiest to see this by thinking about a diagonal matrix. If A is diagonal with elements $a_{11}, a_{22}, ..., a_{nn}$, then volume of the transformed unit cube will be a higher dimensional rectangle and its volume will be $|\text{det}A| = |\prod_{i=1}^n a_{ii}|$. If we multiply any of the columns of A by c, we will scale one of the corners of the rectangle by c, so the volume should be multiplied by c. Similarly, if we add v to the j diagonal element, the volume will be

$$|a_{11} \times ... \times (a_{jj} + v) \times ... \times a_{nn}| = |\prod_{i=1}^n a_{ii}| + |a_{11} \times ... \times v \times ... \times a_{nn}|$$

So, at least for diagonal matrices, if $|\text{det}A|$ is to be the volume of the transformed unit cube, it must be multilinear. It turns out that this is true in general as well.

Okay, so we have defined the determinant in a way that has a nice interpretation, but how can we calculate it? There is an explicit formula for the determinant, but it is not so simple.

Definition 4.2. The **determinant** of a square matrix A is defined recursively as

(1) For 1 by 1 matrices, $\text{det}A = a_{11}$
(2) For n by n matrices,

$$\text{det}A = \sum_{j=1}^n (-1)^{1+j} a_{1j} \text{det}A_{-1,-j}$$

where $A_{-i,-j}$ is the $n-1$ by $n-1$ matrix obtained by deleting the ith row and jth column of A.

$\text{det}A_{-i,-j}$ is called a **minor** of A. $(-1)^{i+j}\text{det}A_{-i,-j}$ is called a **cofactor** of A. The above definition is sometimes called the expansion by cofactors of the determinant. If you are more comfortable with this definition than with the abstract definition (4.1) above, you can consider this formula to be the definition of the determinant. The two defitions are the same, as stated in the following theorem, which we will not prove.

Theorem 4.1. *The two definitions of the determinant, (4.1) and (4.2), are equivalent.*

In some situations, the abstract definition (4.1) is more useful. Proving corollary 4.1 from the abstract definition was not difficult. Proving it using the recursive definition (4.2) is a bit tedious. On the other hand, it will be easier to prove part of lemma 4.2 using the recursive definition than the abstract definition.

Some useful properties of determinants are stated in the following lemma.

Lemma 4.2. *Let A and B be square matrices then*

(1) $\text{det}A^T = \text{det}A$
(2) $\text{det}(AB) = (\text{det}A)(\text{det}B)$
(3) $\text{det}A^{-1} = (\text{det}A)^{-1}$
(4) *Usually,* $\text{det}(A + B) \neq \text{det}A + \text{det}B$
(5) *If A is diagonal,* $\text{det}A = \prod_{i=1}^n a_{ii}$
(6) *If A is upper or lower triangular* $\text{det}A = \prod_{i=1}^n a_{ii}$.

Proof. The proof of all these properties would be quite long, so we will only prove 6. We can just prove it for lower triangular matrices, since $\det A^T = \det A$ and the transpose of upper triangular matrices is lower triangular. We will use induction on the size of A. If A is 1 by 1, the claim is true. If A is n by n, from definition 4.2, we have

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

A is lower triangular, so only the first term in the sum is nonzero.

$$\det A = a_{11} \det A_{-1,-1}$$

$A_{-1,-1}$ is also be lower triangular and is size $n-1$ by $n-1$, so by induction, $\det A = \prod_{i=1}^{n} a_{ii}$. ☐

We can use the determinant to solve systems of linear equations and to calculate A^{-1}.

Theorem 4.2. *Let A be nonsingular. Then,*

(1) $A^{-1} = \frac{1}{\det A} \begin{pmatrix} \det A_{-1,-1} & \cdots & (-1)^{1+n} \det A_{-n,-1} \\ \vdots & \ddots & \vdots \\ (-1)^{1+n} \det A_{-1,-n} & \cdots & (-1)^{n+n} \det A_{-n,-n} \end{pmatrix}$

(2) *(**Cramer's rule**) The unique solution to $Ax = b$ is*

$$x_i = \frac{\det B_i}{\det A}$$

where B_i is the matrix A with the ith column replaced by b.

This theorem can sometimes be useful for analyzing small systems of linear equations and doing comparative statics. For an example, see Simon and Blume 9.3. However, it is a very inefficient way to compute A^{-1} or solve a system of equations.

5. Computational efficiency

In a past year, someone asked about whether there was a more efficient way to calculate the rank of a matrix than Gaussian elimination. The way we usually compare the efficiency of algorithms is by looking at the rate at which number of steps required grows as the size of the problem increases. As an example, we will compare solving a system of equations by Gaussian elimination to using Cramer's rule.

Let's start by considering computing the determinant recursively as in definition 4.2. Let $d(n)$ be the number of steps required to compute the determinant for an n by n matrix. Given the recursive definition, $d(n)$ must satisfy

$$d(n) = nd(n-1) + 2n$$

since $\det A_{-1,-i}$ must be computed n times, and there are n additions and multiplications. We could show by induction that

$$d(n) = 2n! \sum_{k=1}^{n} \frac{1}{(n-k)!}.$$

When comparing algorithms, we often mainly care about the number of steps for large n. We say $d(n)$ is on the order of $f(n)$ or $d(n)$ is big O of $f(n)$ and write $d(n) = O(f(n))$ if $\exists n_0$ such that

$$d(n) \leq Mf(n)$$

for some constant M and all $n \geq n_0$. From the above, $d(n) = O(n!)$. Factorial grows very fast, so the determinant is quite expensive to compute using the recursive definition for large matrices.[11] To actually solve a system of equations using Cramer's rule, we would have to compute $n + 1$ determinants, so Cramer's rule is $O((n+1)!)$.

Now lets look at Gaussian elimination. Let $g(n)$ be the maximal number of steps Gaussian elimination takes on an n by n matrix. In the worst case, Gaussian elimination on the first row involves multiplying the first row by $n-1$ numbers, so there are $n(n-1)$ multiplications, and adding it to each of the $n-1$ rows, so there are $n(n-1)$ additions. The total number of steps to get to the next $n-1$ by $n-1$ submatrix is then $2n(n-1)$. Hence, $g(n)$ is given by

$$g(n) = 2n(n-1) + g(n-1)$$

and,

$$\begin{aligned} g(n) =& 2\sum_{k=1}^{n} k(k-1) \\ =& \frac{2}{3}(n^3 - n) \end{aligned}$$

Therefore, $g(n) = O(n^3)$. To solve a system of equations, we also have to back substitute. The first substitution involes one division, and then $n-1$ additions. The second involves one division and $n-2$ additions. In total there will be $\sum_{k=1}^{n} k = \frac{1}{2}n(n-1)$ steps. So solving a system by Gaussian elimination takes $\frac{2}{3}(n^3 - n) + \frac{1}{2}(n^2 - n) = O(n^3)$ steps. It is much faster than Cramer's rule.

6. Matrix decompositions

There are number of matrix decompositions that are useful for computing solutions to systems of linear equations, inverting matrices, and computing determinants.

Definition 6.1. The **LU decomposition** of a square matrix is a decomposition of the form

$$PA = LU$$

where P is a permutation matrix, L is lower triangular, and U is upper triangular.

We already know how to compute an LU decomposition by performing Gaussian elimination. All row interchange operations can be represented by the permutation matrix P. Adding multiples of one row to another are represented by L, and the resulting row echelon form is U. The LU decomposition exists for any nonsingular matrix. Sometimes it is also required that the diagonal of L by all 1's. In that case, the LU decomposition

[11] Any reasonable computer program will not be computing determinants this way. It will compute some matrix decomposition (LU, QR, or Cholesky, or less likely, SVD) and then use that to compute the determinant.

is unique. The LU decomposition is used to solve linear equations, compute A^{-1}, and compute $\det A$.

Definition 6.2. The **Cholesky decomposition** of a square, symmetric, and positive definite ($\det A > 0$) matrix A is
$$A = LL^T$$
where L is lower triangular with positive entries on the diagonal.

The Cholesky decomposition takes half as many steps to compute than the LU decomposition and can be used for all the same purposes. Symmetric positive definite matrices are quite common in economics. Examples include covariance matrices and the matrix of second derivatives of a convex minimization problem.

Definition 6.3. The **QR decomposition** of an m by n matrix is
$$A = QR$$
where R is m by n and upper triangular and Q is m by m and orthogonal (i.e. $Q^T = Q^{-1}$).

The QR decomposition can be used for all the things the LU decomposition can be used for. The QR decomposition takes twice as many steps to compute, but can be computed more accurately (which can matter for very large matrices, but usually isn't important for small ones). The QR decomposition is also used for solving least square problems. We will (likely) learn how to calculate the QR decomposition and learn an interpretation of Q next lecture.

Definition 6.4. The **singular value decomposition** of a matrix m by n matrix A is
$$A = U\Sigma V$$
where U is m by m and orthogonal, Σ is m by n and diagonal, and V is n by n and orthogonal.

We will study the singular value decomposition in more detail next week.