Elementary Linear Algebra: Math 135A

Lecture Notes
by
Stefan Waner

(2006)

Department of Mathematics, Hofstra University

ELEMENTARY LINEAR ALGEBRA
Math 135 Notes prepared by Stefan Waner

Table of Contents

Text: Elementary Linear Algebra by Howard Anton; 9th ed.

1. Matrix Algebra
(See §§1.3 - 1.4 in text)

Definition 1.1 An **$m \times n$ matrix with entries in the reals** is a rectangular array of real numbers.

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix}$$

Examples 1.2 In class

Notation
(1) If A is any $m \times n$ matrix, then the ijth entry of A will sometimes be written as a_{ij} and sometimes as A_{ij}.
(2) We sometimes write $[a_{ij}]$ or $[A_{ij}]$ to refer to the matrix A. Thus, for example, $[a_{ij}]$ means "the matrix whose ijth entry is a_{ij}." Thus,

$A = [a_{ij}] = [A_{ij}] =$ the matrix whose ijth entry is a_{ij}.

(Similarly, the matrix B is written as $[b_{ij}]$, the matrix Γ as $[\gamma_{ij}]$, etc.)

Definitions 1.2
(i) Two matrices are **equal** if they have the same dimensions and their corresponding entries agree (i.e., they are "the same matrix").
(ii) The $m \times n$ matrix A is **square** if $m = n$. In a square matrix, the entries $a_{11}, a_{22}, a_{33}, \dots, a_{nn}$ form the **leading diagonal** of A.
(iii) A **column vector** is an $n \times 1$ matric for some n; a **row vector** is a $1 \times n$ matrix fro some n.

We now turn to the algebra of matrices.

Definitions 1.3
(a) If $A = [a_{ij}]$ and $B = [b_{ij}]$ are both $m \times n$ matrices, then their **sum** is the matrix

$A + B = [a_{ij} + b_{ij}]$.

In words, "$A+B$ is the matrix whose ijth entry is $a_{ij} + b_{ij}$."
Thus, we obtain the ij$^{\text{th}}$ entry of $A+B$ by adding the ij$^{\text{th}}$ entries of A and B.
(b) The **zero $m \times n$ matrix O** is the $m \times n$ matrix all of whose entries are 0.
(c) If $A = [a_{ij}]$, then $-A$ is the matrix $[-a_{ij}]$.
(d) More generally, if $A = [a_{ij}]$ and if r is a real number, then rA is the matrix $[ra_{ij}]$. This is called **scalar multiplication by *r*.**
(e) If $A = [a_{ij}]$ is any matrix, then its **transpose**, A^T, is the matrix given by

$(A^T)_{ij} = A_{ji}$.

That is, A^T is obtained from A by writing the rows as columns.
(f) The matrix A is **symmetric** if $A^T = A$. In other words, $A_{ji} = A_{ij}$. for every i and j. A is **skew-symmeric** if $A^T = -A$; that is, $A_{ji} = -A_{ij}$ for every i and j.

Examples 1.4 In class

Proposition 1.5 (Algebra of Addition and Scalar Multiplication)
One has, for all real numbers r, s and $m\times n$ matrices A, B, C,

(a) $(A+B)+C = A+(B+C)$ (associativity)
(b) $A+B = B+A$ (commutativity)
(c) $A+O = O+A = A$ (additive identity)
(d) $A+(-A) = (-A)+A = O$ (additive inversion)
(e) $r(A+B) = rA+rB$ (left distributativity)
(f) $(r+s)A = rA + sA$ (right distributativity)
(g) $(rs)A = r(sA)$ (associativity of scalar multiplication)
(h) $1.A = A$ (scalar multiplicative identity)
(i) $0.A = O$ (annihilation by zero)
(j) $(-1).A = -A$ (no name)
(k) $(A+B)^T = A^T + B^T$ (transpose of a sum)

(The last two can actually be deduced from ones above.)

Proof We prove parts (a) and (b) in class, parts (c), (d) and (e) in the homework, and leave the rest as an exercise, parts (c) (d) and (e) are in the homework, while the rest are left as an exercise!

We now consider matrix multiplication.

Definition 1.6 Let $A = [a_{ij}]$ be an $m\times n$ matrix, and let $B = [b_{ij}]$ be an $n\times l$ matrix. Then their *product*, $A\cdot B$ is the $m\times l$ matrix whose ij^{th} entry is given by

$$(A.B)_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \ldots + a_{in}b_{nj} = \sum_{k=1}^{n} a_{ik}b_{kj} \qquad \textbf{Matrix Product}$$

Diagramatically, this entry is obtained by "dotting" the i^{th} row of A with the j^{th} column of B.

Examples in class

Definition 1.7. If $n \geq 1$, then the $n\times n$ identity matrix I_n is the $n\times n$ matrix whose ij^{th} entry is given by

$$\delta_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} .$$

(δ_{ij} is called the Kronecker Delta.) Thus,

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

When there is no danger of confusion, we just write I and drop the subscript n.

Proposition 1.8 (Multiplicative Properties)
Assuming that the sizes of the matrices are such that the operations can be performed, one has for all A, B, C, λ, μ:

(a) $(AB)C = A(BC)$	(associativity)
(b) $A(B+C) = AB + AC$	(left distributativity)
(c) $(A+B)C = AC + BC$	(right distributativity)
(d) $AI = IA = A$	(additive identity)
(e) $\lambda(AB) = (\lambda A)B = A(\lambda B)$	(associativity of scalar multiplication)
(f) $O.A = A.O = O$	(actually follows from (c) and Proposition 1.5.)
(g) $(AB)^T = B^T A^T$	(transpose of a product)

Proof We prove (a) and (d) in class, and assign some of the rest as exercises.
(a) Let $A = [a_{ij}]$, $B = [b_{ij}]$, $C = [c_{ij}]$. Then:

$$
\begin{aligned}
[(AB)C]_{ij} &= \textstyle\sum_k (AB)_{ik} c_{kj} && \text{(by definition of multiplication)} \\
&= \textstyle\sum_k \left(\sum_l a_{il} b_{lk}\right) c_{kj} && \text{(definition of multiplication again)} \\
&= \textstyle\sum_k \left(\sum_l a_{il} b_{lk} c_{kj}\right) && \text{(distributive law in } \mathbb{R}\text{)} \\
&= \textstyle\sum_l \left(\sum_k a_{il} b_{lk} c_{kj}\right) && \text{(commutativity of + in } \mathbb{R}\text{)} \\
&= \textstyle\sum_l a_{il} \left(\sum_k b_{lk} c_{kj}\right) && \text{(distributive law in } \mathbb{R} \text{ applied backwards)} \\
&= \textstyle\sum_l a_{il} (BC)_{lj} && \text{(by definition of multiplication)} \\
&= [A(BC)]_{ij} && \text{(definition of multiplication again)}
\end{aligned}
$$

What seems to be missing is the commutative rule: $AB = BA$. Unfortunately (or perhaps fortunately), however, it fails to hold in general.

Example of failure of commutativity

Examples 1.9
(a) There exist matrices $A \neq O$, $B \neq O$, but with $A.B \neq O$, shown in class. Such matrices A and B are called **zero divisors**.
(b) As a result of (a), the cancellation law fails in general. Indeed, let A and B be as in (i). Then

$A.B = O = A.O,$

but "canceling" the A would yield

$B = O,$

which it isn't.

Exercise Set 1
Anton, Set 1.3 #1–9 odd
Set 1.4 #1(a), (c), 3, 5(a) 7(c), 15, 17

Hand In (Value = 20 points)

1. Prove that if A is any $m\times n$ matrix with the property that $AA^T = O$, then $A = O$.
2. Prove Proposition 1.5(c), (d) and (g).
3. Prove Proposition 1.8(c) and (e).
4. Let

$$A = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}.$$

(a) Determine a simple epxression for A^2 and A^3.
(b) Conjecture the form of the simple expression for A^k, k a positive integer.
(c) Prove your conjecture in **(b)**

2. Systems of Linear Equations

Definition 2.1 A **linear equation** in the n variables $x_1, x_2, \ldots, x_n$ is an equation of the form

$$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b,$$

where the a_i and b are real constants.

Note: This is just an *expression* (i.e. to the computer majors, a string of symbols.) For example, "0 = 1" is an equation, although it is not necessarily true.

Examples 2.2

$3x - y/2 = -11$ is a linear equation in x and y;

$y = x/2 + 67y - z$ can be rearranged to give a linear equation in x, y and z.

$x_1 + x_2 + \ldots + x_n = 1$ is a linear equation in $x_1, x_2, \ldots, x_n$.

$x + 3y^2 = 5$, $3x + 2y - z + xz = 4$, and $y = \sin x$ are not linear.

Definition 2.3 A **solution** of the linear equation

$a_1x_1 + a_2x_2 + \ldots + a_nx_n = b$

is a sequence of numbers $(s_1, s_2, \ldots, s_n)$, each of which is in $\mathbb{R}$, such that, if we replace each x_i by s_i, the expression becomes true. The set consisting of all solutions to the equation is called the **solution set**.

Examples 2.4

(a) $(1, -3)$ is a solution of $3x + y = 0$, since $3(1) + (-3)$ does equal 0.

(b) $(1, 3)$ is not a solution of $3x + y = 0$, since $3(1) + (3)$ does not equal 0.

(c) The solution set of $3x + y = 0$ consists of all pairs $(t, -3t)$, $t \in \mathbb{R}$. We write:

Solution Set = $\{(t, -3t) : t \in \mathbb{R}\}$.

(d) Find the solution set of $x_1 - x_2 + 4x_3 = 6$.

Answer: Choosing x_2 and x_3 arbitrarily, $x_2 = \alpha$, $x_3 = \beta$, say, we can now solve for x_1, getting $x_1 = 6+\alpha-4\beta$. Thus, solutions are all triples of the form $(6+\alpha-4\beta, \alpha, \beta)$, whence

Solution Set = $\{ (6+\alpha-4\beta, \alpha, \beta) : \alpha, \beta \in \mathbb{R}\}$.

(e) The solution set of the linear equation $0x - 0y = 1$ is empty , since there are no solutions. Thus,

Solution Set = Ø

in this case.

Definitions 2.5 A system of linear equations in the variables $x_1, x_2, \ldots, x_n$ is a finite set of linear equations in $x_1, x_2, \ldots, x_n$. Thus, a system of m equations in $x_1, x_2, \ldots, x_n$ has the form

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots & \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n &= b_m \end{aligned}$$

A **solution** to a system of equations is a sequence of numbers $(s_1, s_2, \ldots, s_n)$ each of which is in $\mathbb{R}$ such that, if we replace each x_i by s_i, all the expressions become true. As before, the set consisting of all solutions to the system of equations is called the solution set.

Examples 2.6
(a) The system

$$4x_1 - x_2 + 3x_3 = -1;$$
$$3x_1 + x_2 + 9x_3 = -4$$

has a solution $(1, 2, -1)$. However, $(1, 8, 1)$ is not a solution, since it only satisfies the first equation.
(b) **Scholium** The system

$$x + y = 1$$
$$2x + 2y = 3$$

has no solutions.

Proof Suppose there is a solution, say (s_1, s_2). Then we would have:

$s_1 + s_2 = 1$ *as a real number,* and
$2s_1 + 2s_2 = 3$ " " " ".

But then, the second equation says that $s_1 + s_2 = 3/2$.

Thus we have:

$s_1 + s_2 = 1$ as a real number, and
$s_1 + s_2 = 3/2$ " " " ".

Subtracting, we get $0 = 1/2$. But this is absurd!!
Thus our assumption (that there was a solution) was false. That is, there is no solution. ❐
Remark This is called a *proof by contradiction*. You assume the negation of what you are trying to prove, and then deduce a contradiction, e.g. $0 = 1$.
(c) Geometric interpretation of systems with two unknowns, and the possible outcomes—discussed in class.

Definition 2.7 A system of equations which has at least one solution is called **consistent**. Otherwise, it is called **inconsistent**. Thus, for example, 2.6(a) is consistent, while 2.6(b) is inconsistent.

Matrix Form of System of Equations
Consider the matrix equation

$$AX = B,$$

where $A = [a_{ij}]$ is an $m\times n$ matrix, and where X is $n\times 1$, so that B must be $m\times 1$. Diagramatically, we have

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ \dots \\ x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}$$

Since the left hand side, when multiplied out, gives a $m\times 1$ matrix, we have the following equation of $m\times 1$ matrices:

$$\begin{bmatrix} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \end{bmatrix} = \begin{bmatrix} b_1 \\ b_2 \\ \dots \\ b_m \end{bmatrix}$$

But, for these matrices to be equal, their corresponding entries must agree. That is, we get the following (rather familiar) system of m linear equations in n unknowns.

$$\left.\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{array}\right\} \dots\dots (*)$$

Thus, we can represent the system (*) by the matrix equation $AX = B$, where X is the column matrix of unknowns, and A is the **coefficient matrix**. Now, to solve for the matrix X, it seems that all we need do is multiply both sides by A^{-1}, getting $X = A^{-1}B$. But this is no simple matter! For starters, we don't yet have a way of getting A^{-1}, and in fact it may not even *exist*. Indeed, inversion of matrices is quite a business, and concepts related to inversion will form the bulk of the rest of this course.

Example in class

Instead, we'll use a different approach to solve a system:

Definition 2.8 If

$$\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2 \\ \dots\dots\dots\dots\dots\dots\dots \\ a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m \end{array}$$

is a system of linear equations, then the $m\times(n+1)$ matrix

$$\left[\begin{array}{ccccc|c} a_{11} & a_{12} & a_{13} & \dots & a_{1n} & b_1 \\ a_{21} & a_{22} & a_{23} & \dots & a_{2n} & b_2 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & a_{m3} & \dots & a_{mn} & b_m \end{array}\right]$$

is called the **augmented matrix** of the system.

Example 2.9 The augmented matrix for the system

$$\begin{array}{rcrcrcl} 2x_1 & - & 4x_2 & + & x_3/2 & = & -1 \\ x_1 & & & - & 11x_3 & = & 0 \\ & & x_2 & + & x_3/2 & = & 3/5 \end{array}$$

We find solutions to systems by replacing them by "equivalent" systems, obtained by the following operations.

Multiply an equation by a constant
Interchange two equations
Add a multiple of one equation to another.

Notes

1. By the laws of arithmetic, if $(s_1,\ s_2,\ ...,\ s_n)$ is a solution to a linear system, then it remains a solution to the system obtained by performing any of the above operations.
2. We don't lose any information by these operations; we can always recover the original set of equations by doing the reverse operations—e.g. the reverse of adding twice equation 3 to equation 1 is subtracting twice equation 3 from equation 1.
3. Since the operations are reversible, we can use Note (1) above backwards; a solution to a system of equations after an operation is also a solution to the original system. In other words, *the solution set is not changed by operations of this type*.

Since a system of equations is "completely represented" by the augmented matrix, we make the following definition.

Definition 2.10 An **elementary row operation** is one of the following:

1. Multiplication of a row by a nonzero* constant.
2. The interchanging of two rows.
3. Addition of a multiple of one row to another.

They are called *elementary* because any kind of legitimate manipulation you do may be obtained by doing a sequence of elementary row operations. Thus they are the "building blocks" for getting "equivalent" matrices.

By the notes before Definition 2.10, we have the following.

Proposition 2.11 (Elementary Row Operations Do Not Effect Solution Set)
Let B be obtained from the augmented matrix A of a system of linear equations by performing a finite sequence of elementary row operations. Then the systems of linear equations represented by A and B have the same solution set..

Example of use of elementary row operations to solve a system (p. 54 in Kolman)

Definition 2.12. A matrix is in **row-reduced echelon form** if:

1. The first nonzero entry in each row (the leading entry) is a 1.

---

* If you multiply by zero, you've lost the corresponding equation.

2. Each column which contains a leading entry has zeros everywhere else.
3. The leading entry of any row is further to the right than those of the rows above it.
4. The rows consisting entirely of zeros are at the bottom.

Examples 2.13

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

$\begin{bmatrix} 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ is not.

Examples 2.13 Solve the systems represented by **(a)**, **(b)** and **(c)** in 2.13.

Illustration of procedure to row-reduce an arbitrary matrix.

Exercise Set 2

Anton §1.1, #1, 3, 7.

§1.2 # 1, 5, 7 (use our method for each of these), 13, 25

Web-site For on-line pivoting go to the URL

www.finitemath.com → Online Utilities → Pivot and Gauss-Jordan Tool

Use the web-site to do some of the above problems.

Hand In (Value = 20 points)

1 (a) Show that the system

$x + y = 1$

$0y = 1$

has no solutions.

(b) Show that the system

$x + y = 1$

$x - y = 1$

has a unique (i.e. exactly one) solution. [Hint: look at the way we handled the Scholium.]

(c) Show analytically that any system of the form

$ax + by = c$

$\lambda ax + \lambda by = \lambda c \quad (a, b, c, \lambda \in \mathbb{R}, a \neq 0)$

has infinitely many solutions, and write down its solution set.

2. Give an example of a row operation e with the property that, if B is obtained from A by performing the operation e:

(a) Every solution to a system represented by A, remains a solution to the system represented by B.

(b) There are matrices A such that not every solution of the system represetned by B is a solution to the system represented by A.

3. Prove that if a *square* matrix A does not row-reduce to the identity, then it row-reduces to a matrix with a row of zeros (meaning one or more rows of zeros).

4. Show how the interchange of two rows can be accomplished by a sequence operations of the other types 1 and 3 in which we only use type 1 once and multiply by -1.

☆ **Considerable Extra Credit** If, without the help of others in the class you can prove—and verbally defend your proof—that the row-reduced form of every matrix is unique, 10 points will be added to your final examination score. Deadline: one month before finals week.

3. Homogeneous Systems[†]

Just as homogeneous differential equations are important tools in solving them, so one has a notion of "homogeneous" systems of linear equations.

The matrix form of systems of equations makes it easy to establish many results. For example:

Definition 3.1 A system of linear equations is said to be **homogeneous** if all the constant terms are zero. That is, the system takes the form

$$\boxed{\begin{array}{c} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = 0 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = 0 \\ \cdots\cdots\cdots\cdots\cdots\cdots\cdots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = 0 \end{array}}$$

Homogeneous System

Recall from Section 2 that every system of linear equations can be represented in the matrix form

$AX = B$,

where A is the coefficient matrix, X is the column matrix of unknowns, and B is a given column matrix. Here, we have $B = 0$, so a homogeneous system can be represented as follows.

$\boxed{AX = O}$ **Homogeneous System: Matrix Form**

Example in class

Remark (0,0, . . . ,0) is a solution of *every* homogeneous system, so they are always consistent. The zero solution is called the *trivial solution,* while any other solutions are called *nontrivial*.

Lemma 3.2 (Linear Combinations of Solutions to a Homogeneous System)
If $X_1 = (s_1, s_2, \ldots, s_n)$ and $X_2 = (t_1, t_2, \ldots, t_n)$ are solutions to a homogeneous system, then so is $\alpha X_1 + \beta X_2 = (\alpha s_1+\beta t_1, \alpha s_2+\beta t_2, \ldots, \alpha s_n+\beta t_n)$ for any real numbers α and β.

(Compare the situation with homogeneous solutions of diff. eqns.)
Proof in class; using matrix form.

Remarks
1. In particular, the lemma implies that the "sum" and "difference" of any two solutions is also a solution.
2. We refer to $\alpha X_1 + \beta X_2$ as a **linear combination** of X_1 and X_2. Similarly, $\alpha X_1 + \beta X_2 + \gamma X_3$ is a linear combination of X_1, X_2, and X_3.
3. Lemma 3.2 says that every solution to a homogeneous system can be represented as a linear combination of a collection of specific solutions $X_1, X_2, \ldots, X_r$. We refer to a collection of vectors as **linearly independent** if none of them is a linear combination of the others.

[†] Anton's treatment of homogenous systems is sketchy, and scattered all over the place.

Proposition 3.3 (Solutions of Homogeneous Systems)
Any homogeneous system of equations either has exactly one solution (the trivial one), or else it has infinitely many solutions.

Proof in class.

Example 3.4 Find the solution set of the homogeneous system, and also a "basis" for the span.

$$\begin{aligned}
2x_1 + 2x_2 - x_3 \phantom{{}- 3x_4} + x_5 &= 0\\
-x_1 - x_2 + 2x_3 - 3x_4 + x_5 &= 0\\
x_1 + x_2 - 2x_3 \phantom{{}- 3x_4} - x_5 &= 0\\
x_3 + x_4 + x_5 &= 0
\end{aligned}$$

The above example illustrates the following.

Proposition 3.5 A homogeneous system with fewer equations than unknowns always has infinitely many solutions.

Proof Let the given system have m equations in n unknowns, with $m < n$. After row reduction, we end up with $\leq m < n$ equations of the form:

$$x_i - stuff = 0$$

i.e., $x_i = -stuff$.

Thus we have no free choices for these variables. However, the remaining ones, (and there *are* remaining ones, since $m < n$), can now be chosen freely, so we can choose some $\neq 0$, getting a nontrivial solution. By Proposition 3.3, we are therefore guaranteed infinitely many solutions. ◆

Now we apply these results to non-homogeneous systems:

Definition 3.6 Given any system of equations, the **associated homogeneous system** is obtained by replacing all the constants on the right-hand-side by zero. In other words, the homogeneous system associated with $AX = B$ is $AX = O$.

Lemma 3.7 If X_1 and X_2 are solutions of the (not necessarily homogeneous) system $AX = B$, then $X_1 - X_2$ is a solution to the associated homogeneous system $AX = O$.

Proof in class. ✱

Proposition 3.8 (Form of Solution to General System)
Let P be any particular solution of the system $AX = B$. Then *every* solution to the system $AX = B$ has the form

$$X = P + H,$$

where H is a solution to the associated homogeneous system $AX = O$.

Proof in class. ✱

Proposition 3.9 (Solutions of General Linear Systems)
Every system of linear equations (homogeneous or not), has either no solutions, exactly one solution, or infinitely many solutions.

Proof in Exercise Set 3.

Exercise Set 3
§1.2 # 7, 13 (solve the associated homogeneous equations—you did the non-homogeneous ones in the preceding set—and verify Proposition 3.8 in each case)
Hand In: (Value = 25 points)
1 (a) Give an example to show that Lemma 3.2 fails in the case of non-homogeneous systems of equations.
(b) Give an example to show that Proposition 3.5 fails in the case of non-homogeneous systems of equations
2. Prove Proposition 3.9.
3. A certain nonhomogeneous system has the particular solution $(-1, 0, 1)$,
and the augmented matrix of the associated homogeneous system is

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

Find the solution set.
4. Let A be an $n\times n$ matrix . Prove that the matrix equation $AX = O$ has a non-zero solution if and only if the row-reduced form of the matrix A has a row of zeros. [You might find Exercise Set 2 #3 as well as the method of proof in Proposition 3.5 useful in your argument.]
5. Now modify your proof of Exercise 4 to show:
Let A be an $n\times n$ matrix . Prove that the matrix equation $AX = B$ has a unqiue solution for every column matrix B if and only if A row reduces to the identity.

---

4. Relations and Row-Equivalence

We pause now to do a little abstract algebra. Let A be any set. Let $A\times A$ be the set of all pairs (a, a') with a and a' in A. E.g. $\mathbb{R}\times\mathbb{R}$ is the set of all ordered pairs of real numbers = the set of "points in the plane."

Examples of Cartesian products, esp. $\mathbb{Z}\times\mathbb{Z}$.

Definition 4.1 A relation R on the set A is a specified collection R of pairs in A. If the pair (a, b) is in R, we write aRb, or $a \sim_R B$ (or simply $a \sim b$), and say that ***a* stands in the relation *R* to *b***.

Examples 4.2

(a) Define a relation < on $\mathbb{R}$ by $x < y$ if x is less than y. Then $1 < 2$, $-4 < 11$, but $2 \nless 1$.
(b) Geometric illustration of this on $\mathbb{Z}\times\mathbb{Z}$.
(c) Define a relation D on $\mathbb{R}$ by $x \sim_D y$ if $y = 2x$. Then $1 \sim 2$, $2 \sim 4$, $0 \sim 0$, $-1 \sim -2$, but $2 \nsim 1$.
Note that, as a set, $D = \{(x,y) \in \mathbb{R}^2 : y = 2x\}$, and can be represented by the points on a straight line in $\mathbb{R}^2$.
(d) Let M(m,n) be the set of all $m\times n$ matrices. Define a relation ≈ on M(m,n) by $A \approx B$ if B can be obtained from A using a finite sequence of elementary row operations. Then, e.g.

$$\begin{bmatrix} 1 & 0 & 4 \\ 3 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \approx \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

but

$$\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \not\approx \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

We refer to the relation ≈ as **row equivalence**.
(e) Plane Geometry The relation of **congruence** is defined as follows: If $A$and B are subsets of the plane, then we define $A \equiv B$ to mean that B can be obtained form A using a finite sequence of rotations and translations.

Definition 4.3 An **equivalence relation** on the set A is a relation ≈ on A with the following properties:

(i) If a ≈ b then b ≈ a (symmetric)
(ii) For all $a \in A$, one has $a \approx a$ (reflexive)
(iii) If $a \approx b$ and $b \approx c$, then $a \approx$ c (transitive).

Examples 4.4

(a) Define the relation $\equiv$ on the integers $\mathbb{Z}$ by $n \equiv m$ if $n - m$ is a multiple of 2 (that is. either both are even, or both are odd). Then:

Proposition ≡ is an equivalence relation.
Proof Checking the three requirements:
(i) If $n \equiv m$, then $n-m$ is even. Thus $m-n = -(n-m)$ is even also.
Thus, $m \equiv n$.
(ii) Since $n-n = 0$ is even, one has $n \equiv n$.

(iii) If $\underline{n \equiv m}$ and $\underline{m \equiv l}$, then $n-m$ and $m-l$ are both even.

Adding, $n-l = n-m + m-l =$ even + even = even.

Thus, $\underline{n \equiv l}$.

(Underlined parts give the definition.)▮

This equivalence relation is called **equivalence mod 2.**

(b) The relation D of Example 4.2(c) is *not* an equivalence relation because $1 \sim 2$ whereas $2 \not\sim 1$, so it fails to be reflexive.

Proposition 4.5 Row equivalence ≈ is an equivalence relation.

Proof. We first make an observation.

For every elementary row operation, e, there is a row operation e' which undoes the operation e. Indeed, if e is:

multiplication of a row by λ, we take e' to be multiplication by λ^{-1}.

interchanging two rows, we take $e' = e$

replacing Row$i$ by Row$i$ + λRow$j$, we take e' as replacing Rowi by Row$i$ − λRow$j$.

We now check the three requirements for an equivalence relation.

(i) Symmetry: <u>If $A \approx B$, then</u> B is obtained from A using a sequence of elementary row operations. Applying the corresponding inverse operations e' in reverse to B now gives us A back. Thus $\underline{B \approx A}$.

(ii) Since A is obtained from A by multiplying Row1 by 1, we conclude that $\underline{A \approx A}$ for every m¿n matrix A.

(iii) Transitivity: <u>If $A \approx B$ and $B \approx C$, then</u> this means that we can get C by first applying the necessary elementary row operations to A that yield B, and then applying those that yield C (from B). Thus $\underline{A \approx C}$. ✦

Theorem 4.6 Any matrix $A \in$ M(m,n) is row-equivalent to a matrix in row reduced echelon form.

(This says that you can row-reduce any matrix in sight.)

Proof A little messy, but will be discussed in class. ✲

Exercise Set 4

Investigate which of the three properties (<u>S</u>ymmetry, <u>R</u>eflexivity, <u>T</u>ransitivity) hold for the following relations on M(n,n).

A. $A \sim_C B$ if A is *column-equivalent* to B; that is, B can be obtained from A by applying a finite sequence of elementary column operations (defined in the same way as row operations).

B. A❊B if $A = 2B$.

C. $A \nabla B$ if $A-B$ is upper triangular (zero below the leading diagonal).

D. $A \Delta B$ if $AB = 0$.

E. A✯B if $AB = BA$.

F. A▮B if B can be obtained from A by applying a *single* elementary row operation.

[Answers: **A.**S, R, T **B.** none **C.** S, R **D.** none **E.** S, R **F.** S, R]

Hand In: (Value = 20 points)

1. Give an example of a relation on $\mathbb{Z}$ which is

  (i) Transitive but not reflexive

  (ii) Neither transitive nor reflexive

  (iii) Transitive, reflexive, but not symmetric.

2. Prove or disprove the following assertions about relations on $\mathbb{Z}$:

  (i) The relation $x \sim y$ if $x \neq y$ is an equivalence relation.

  (ii) The relation $x \sim y$ if $x - y$ is a multiple of 3 is an equivalence relation.

3. Assume that $ad - bc \neq 0$. Show that the matrix

  $\begin{bmatrix} a & b \\ c & d \end{bmatrix}$ is row-equivalent to $\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$.

4. Show that the following statements are equivalent for a *square* mattrix A:

  (a) A is row-equivalent to the identity matrix.

  (b) A is not row-equivalent a matrix with a row of zeros.

[Hint for **(a)** $\Rightarrow$ **(b)**: Exercise Set 3 #4]

5. Inverses and Elementary Matrices

Definition 5.1 An $n\times n$ matrix A is said to be **invertible** if there is an $n\times n$ matrix B such that $AB = BA = I$. We then call B an **inverse of A.** If A has no inverse, then we say that A is **singular.**

Proposition 5.2 An invertible matrix A has exactly one inverse, (and we therefore refer to it as A^{-1}, so that $AA^{-1} = A^{-1}A = I$).

Proof There are many ways to prove this. S'pose B and C were both inverses of A. Then one has

$B = (CA)B = C(AB) = C$ ❇

Example 5.3 The inverse of any 2×2 matrix is given by the formula:

$$\boxed{\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}}$$

Inverse of 2×2 Matrix

If $ad-bc = 0$, then in fact the inverse does not exist (the matrix is singular) as we shall find out later. (Compare Exercise Set 2 #3).

Proposition 5.4 (Properties of Inverses)
Assume A &B are invertible. Then:

- **(a)** AB is invertible, and $(AB)^{-1} = B^{-1}A^{-1}$.
- **(b)** A^{-1} is invertible, and $(A^{-1})^{-1} = A$.
- **(c)** If $\lambda \neq 0$, then λA is invertible, and $(\lambda A)^{-1} = \frac{1}{\lambda} A^{-1}$.
- **(d)** If A is invertible, then so is A^T, and $(A^T)^{-1} = (A^{-1})^T$.
- **(e)** If r and $s \in \mathbb{Z}$, then $A^rA^s = A^{r+s}$.
  (**Defn.** $A^{-n} = (A^{-1})^n$ for $n > 0$.)

Proof We do (a) in class, (b) & (c) & (d) appear in homework, and (e) will be outlined in class ❇

Remark It follows from part (a), that, if $A_1, A_2, \ldots, A_r$ are all invertible matrices, then so is $A_1A_2\ldots A_r$, and its inverse is $A_r^{-1}A_{r-1}^{-1}\ldots A_2^{-1}A_1^{-1}$.

Definition 5.5 An $n\times n$ matrix E is called an **elementary** matrix if it can be obtained from the $n\times n$ identity I using a single elementary row operation. (See Definition 2.10.) (If you need a fancy row operation, such as replacing row r by 6(row r) + 5(row s), it isn't elementary.)

It follows that elementary matrices have the form:

(i) The identity with one of the diagonal 1's replaced by k;
(ii) The identity with two of its rows interchanged;
(iii) The identity with a single off-diagonal entry k.

Elementary Matrices

Examples in class

Lemma 5.6 If the elementary matrix E is obtained from I by doing an elementary row operation e, and if A is any $n \times n$ matrix, then $E.A$ is obtained from A by doing the same operation e to A. In other words, if

$$E = e(I),$$

then $\quad EA = e(A)$

for every matrix A (such that EA is defined).

Proof Since elementary row operations are of one of three types listed on page 3, with corresponding elementary matrices listed above, it suffices to check these three cases. This we do in class. ❇

Example of how it works in class

Lemma 5.7 Every elementary matrix is invertible, and its inverse is also an elementary matrix.

Proof in class. ❣

Theorem 5.8 (Invertible Matrices and Systems)

The following statements are equivalent for an $n \times n$ matrix A.

(a) A is invertible.

(b) For every B, the system $AX = B$ has exactly one solution.

(c) $AX = O$ has only the trivial solution.

(d) A is not row-equivalent to a matrix with a row of zeros.

(e) A is row-equivalent to the identity matrix I.

(f) A is a product of elementary matrices.

Proof We use a "circle of implications."

(a)⇒(b):

Assume A is invertible. Then A^{-1} exists, and it can be checked by substitution that $X = A^{-1}B$ is a solution of $AX = B$. we must show (b). But, given $AX = B$, we can multiply both sides by A^{-1}, getting $A^{-1}AX = A^{-1}B$ i.e.., $X = A^{-1}B$, showing that the solution X is unique. That it is a solution can be seen by substitution.

(b)⇒(c):

Assuming (b), Exercise set 3 #5 says that the matrix A row reduces to the identity. The same exercise, with $B = O$ now tells us that $AX = O$ must have a unique solution. Since the trivial solution is always a solution, that must be the unique solution.

(c)⇒(d):

This is Exercise Set 3 #4 (in negation form).

(d)⇒(e):

This is Exercsie Set 4 #4.

(e)⇒(f):

Assume that A row-reduces to I. Then there is a sequence of row operations $e_1, e_2, \ldots, e_p$ such that application of e_1 then $e_2, \ldots,$ then e_p to A results in I. Thus:

$$e_p(e_{p-1}(\ \ldots (e_2(e_1(A)))\ \ldots\ )) = I.$$

By Lemma 5.6, if we take E_j to be the elementary matrix corresponding to e_j, we now have

$$E_p(E_{p-1}(\ \ldots (E_2(E_1 \cdot A))\ \ldots\ )) = I,$$

that is,

$$E_p.E_{p-1}.\ \ldots\ E_2E_1.A = I.$$

Now multiply both sides by $E_1^{-1}E_2^{-1}\ \ldots\ E_{p-1}^{-1}E_p^{-1}$ on the left. This yields

$$A = E_1^{-1}E_2^{-1}\ \ldots\ E_{p-1}^{-1}E_p^{-1}.$$

Since, by Lemma 5.7, each E_j^{-1} is an elementary matrix, we have shown that A is a product of elementary matrices.

(d)⇒(a):

This follows from the remark following Proposition 5.4: Products of invertible matrices are invertible. ❷

Remark S'pose A has all its entries whole numbers, and suppose that A is invertible with A^{-1} also having integral entries. Then, just as in the theorem, it can be shown that A is a product of *integer* elementary matrices. In the context of integer matrices, the following question is the subject of ongoing research: Find the minimum number of elementary matrices of which A is a product. I believe that a colleague, David Carter, has shown that in the 2×2 case, the answer is five. (It is not known for larger matrices.)

Remark 5.8 With the notation in the proof of the theorem, s'pose A reduces to I, so that

$$e_p(e_{p-1}(\ \ldots (e_2(e_1(A)))\ \ldots\ )) = I.$$

Then, as in that part of the proof,

$$E_p.E_{p-1}.\ \ldots\ E_2E_1.A = I,$$

so that

$$\begin{aligned} A^{-1} &= E_p.E_{p-1}.\ \ldots\ E_2E_1 \\ &= E_p.E_{p-1}.\ \ldots\ E_2E_1I \\ &= e_p(e_{p-1}(\ \ldots (e_2(e_1(A)))\ \ldots\ )). \end{aligned}$$

Thus, A^{-1} is obtained by applying the same sequence of row operations to I that were used to reduce A^{-1} to the identity. Thus we have a method for finding the inverse of *any* invertible matrix.

Example in class

Exercise Set 5
Anton §1.4 #7, 11, 13, 18
Hand In (Value = 25)
1. Prove Proposition 5.4 (b), (c), (d).
2. Find a non-zero, non-identity 2×2 matrix E such that $E^2 = E$, and conclude that E is a zero divisor (that is, there exists a matrix F, other than O, such that $EF = O$).
3. Show that the inverse of an upper triangular matrix ($A_{ij} = 0$ if $i > j$) is upper triangular.
4. Express the matrix

$$A = \begin{bmatrix} a & b & c \\ 0 & d & e \\ 0 & 0 & f \end{bmatrix} \quad (a,\ d,\ f \text{ all non-zero})$$

as a product of six elementary matrices.
5. Let A be any $n\times n$ matrix. Show that every system $AX = B$ has a solution if and only if every system $AX = B$ has a *unique* solution.

6. Determinants

Definition 6.1 A **permutation** σ of the n letters $\{1, 2, \ldots, n\}$ is a sequence $(\sigma(1), \sigma(2), \ldots, \sigma(n))$ in which each of the letters $1, \ldots, n$ occurs exactly once.

Examples 6.2
(a) (4, 2, 1, 3) is a permutation of {1, 2, 3, 4}, and here,
$\sigma(1) = 4,\ \sigma(2) = 2,\ \sigma(3) = 1,\ \sigma(4) = 3$.
(b) $(1, 2, \ldots, n)$ is a permutation of $\{1, 2, \ldots, n\}$, with $\sigma(i) = i$ for all i. This permutation is called the **identity permutation**.
(c) (1, 3, 3, 5, 2) is *not* a permutation of {1, 2, 3, 4, 5}, since 3 repeats and 4 is missing.

Remarks 6.3
(a) The number of permutations of $\{1, 2, \ldots, n\}$ is $n(n-1)\ldots 2.\ 1 = n!$
(b) As the notation $\sigma(i)$ in the examples suggests, we may view a permutation σ as a function that assigns to each member of $\{1,2, \ldots, n\}$ another (possibly the same) member of $\{1,2, \ldots, n\}$. Further, such a function σ is a 1-1 correspondence. We write

$$\sigma\colon \{1, 2, \ldots, n\} \longrightarrow \{1, 2, \ldots, n\}$$

to indicate that σ is a function from $\{1, 2, \ldots, n\}$ to $\{1, 2, \ldots, n\}$.

Cycle Representation and Decomposition into 2-cycles
In class

Definition 6.4 Given a permutation σ, an **inversion** is a pair of integers $i <$ j such that $\sigma(i) > \sigma(j)$. A permutation σ of $\{1,2, \ldots, n\}$ is called an **even permutation** if the total number of inversions is even. Otherwise, σ is called an odd permutation. If σ is an even permutation, we write $sgn(\sigma) = +1$, else we write $\mathrm{sgn}(\sigma) = -1$ (for σ an odd permutation).

Geometric Interpretation of Parity Note that if one arranges the elements of $\{1, 2, \ldots, n\}$ down two columns and represents σ by a sequence of arrows, one gets a crossing for each inversion. Thus σ is odd if there is an odd number of crossings.

Remark In the extra credit homework, you will show that sgn of a transposition is -1, and that $\text{sgn}(\sigma\circ\mu) = \text{sgn}(\sigma)\text{sgn}(\mu)$ for every pair of permutations σ, μ.

Example 6.5
The transposition $(2, 5)$ is odd, since there are 5 inversions; or, by the extra credit stuff, it is a product of a single transposition.
The identity permutation is always even, as there are 0 inversions.

Now let

$$A = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \ldots & a_{1n} \\ a_{21} & a_{22} & a_{23} & \ldots & a_{2n} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ a_{n1} & a_{n2} & a_{n3} & \ldots & a_{nn} \end{bmatrix}.$$

be any square matrix, and let ß be any permutation of $\{1, 2, \ldots, n\}$. Then define $\Pi_\sigma(A)$ to be the number

$$\Pi_\sigma(A) = a_{1,\sigma(1)}\, a_{2,\sigma(2)} \, . \, . \, . \, a_{n,\, \sigma(n)} \, .$$

Thus $\Pi_\sigma(A)$ is a product containing exactly one entry from each row.

Example If

$$A = \begin{bmatrix} 2 & 6 & 9 \\ 1 & -1 & 3 \\ 0 & -1 & 8 \end{bmatrix} \text{ and } \sigma = (3,\ 1,\ 2), \text{ then}$$

$$\Pi_\sigma(A) = 9\times1\times(-1) = -9.$$

Note that the number of possible expressions $\Pi_\sigma(A)$ is $n!$, since this is the number of possible permutations of n letters.

Definition 6.6 We define the determinant of the $n\times n$ matrix A by the formula:

$$\det(A) = \Sigma_\sigma \Pi_\sigma(A) \,\text{sgn}(\sigma),$$

where the sum is taken over all permutations σ of $\{1, 2, \ldots, n\}$. (Thus there are $n!$ terms we are adding together).

Examples 6.7

(a) Let $A = \begin{bmatrix} 2 & 6 & 9 \\ 1 & -1 & 3 \\ 0 & -1 & 8 \end{bmatrix}$. We must now add together 9 terms:

σ	sgn(σ)	$\Pi_\sigma(A)$sgn(σ)
(1, 2, 3)	+1	$+(2\times-1\times8)$
(1, 3, 2)	−1	$-(2\times3\times-1)$
(2, 3, 1)	+1	$+(6\times3\times0)$
(2, 1, 3)	−1	$-(6\times1\times8)$
(3, 1, 2)	+1	$+(9\times1\times-1)$
(3, 2, 1)	−1	$-(9\times-1\times0)$

Note that this can be illustrated with diagonal lines (in class).
Thus, to get det(A), we add everything in the right-hand column, getting

$$\det(A) = -16 + 6 - 48 - 9 = -67 \text{ (I think).}$$

(b) For a general 2×2 matrix, there are only two permutations; the even permutation (1,2) and the odd permutation (2,1).

Thus, $\det(A) = a_{11}a_{22} - a_{12}a_{21}$.

Notation If A has rows $r_1, r_2, \ldots, r_n$, we shall write

$$A = [r_1, r_2, \ldots, r_n].$$

Lemma 6.8
(a) If A is any square matrix containing a row of zeros, then $\det(A) = 0$.
(b) One has, for any i ($1 \le i \le n$),

$$\det[r_1, \ldots, r_i+s_i, \ldots, r_n] = \det[r_1, \ldots, r_i, \ldots, r_n] + \det[r_1, \ldots, s_i, \ldots, r_n]$$

(c) If two rows A are the same, then $\det(A) = 0$.

Proof
(a) If one has a row of zeros, then each $\Pi_\sigma(A)$ is zero.
(b) If $\sigma \in \Sigma_n$, then

$$\Pi_\sigma[r_1, \ldots, r_i+s_i, \ldots, r_n] = \Pi_\sigma[r, \ldots, r_i, \ldots, r_n] + \Pi_\sigma[r_1, \ldots, s_i, \ldots, r_n]$$

by the distributive law in $\mathbb{R}$. Summing over $\sigma \in \Sigma_n$ now gives the result.
(c) S'pose rows i and j are the same. If $\sigma \in \Sigma_n$ with $\sigma(i) = p$ and $\sigma(j) = $ q. This means that in the summand of the determinant coming from σ, we have the factor $A_{ip}A_{jq}$. Now let σ' be the composition of σ with the permutation (i, j); that is,. $\sigma'(x) = \sigma\circ(i,j)(x)$, so that

$$\sigma'(i) = \sigma\circ(i,j)(i) = \sigma(j) = q$$

and similarly $\sigma'(j) = p$. Note that σ and σ' agree on every other element. By the extra credit problem, $\text{sgn}(\sigma') = -\text{sgn}(\sigma)$, so that the perumation σ' contrubutes the same product $A_{iq}A_{jp} = A_{ip}A_{jq}$ (since the ith and j th rows are the same) but with opposite sign. Therefore, all the summands in det(A) cancel in paris, and we get 0.
✳

We consider the effect of row operations on determinants. (See pp. 95-97 of Kolman.)

Proposition 6.9 (Effects of Row Operations on the Determinant)
Let A' be obtained from A by a row operation of the type e. then:
(a) If e is multiplication of some row by λ, then $\det(A') = \lambda\det(A)$.
(b) If e is addition of μ times one row to another, then $\det(A') = \det(A)$.
(c) If e interchanges two rows, then $\det(A') = -\det(A)$.

Proof
(a) Here, $\Pi_\sigma(A') = \lambda\Pi_\sigma(\mathrm{A})$ for each $\sigma \in \Sigma_n$. Thus,

$$\det(A') = \Sigma_\sigma\Pi_\sigma(A')\mathrm{sgn}(\sigma) = \Sigma_\sigma\lambda\Pi_\sigma(A)\mathrm{sgn}(\sigma) = \lambda\Sigma_\sigma\Pi_\sigma(A)\mathrm{sgn}(\sigma) = \lambda \det(A).$$

(b) S'pose A' is obtained from A by replacing r_i by $r_i + \lambda r_j$. Then, by the lemma (part (b)), one has

$$\begin{aligned}\det(A') &= \det[r_1, \ldots, r_i+\lambda r_j, \ldots, r_n] \\ &= \det[r_1, \ldots, r_i, \ldots, r_n] + \det[r_1, \ldots, \lambda r_j, \ldots, r_n] \\ &= \det(A) + \lambda\det[r_1, \ldots, r_j, \ldots, r_n],\end{aligned}$$

by part (a). Note that the second term has row j appearing twice; once in the j^{th} slot and once in the i^{th} slot. Thus it vanishes by Lemma 6.8(c). Hence $\det(A') = \det(A)$, as required.
(c) Recall that any row swap can be done using a sequence of operations of type (b) combined with and a single multiplication of a row by -1 (see Exercise Set 2 #4).
❈

Remarks
1. It follows from the above that we can calculate the determinant of any (square) matrix by keeping track of the row operations required to reduce it. In fact, we see that we need only keep track of the number of times we switch two rows as well as the numbers by which we multiply rows. If the matrix does not reduce to I, then the determinant vanishes.
2. It follows from part (b) of 6.9 that, if A has two identical rows, then $\det(A) = 0$. More generally, it follows from the above remark that if A does not row-reduce to the identity, then $\det(A) = 0$.

Lemma 6.10 If A is any square matrix in row echelon form, one has $\det(A) = 1$ if $A = \mathrm{I}$, and $\det(A) = 0$ otherwise.

The proof is in Exercise Set 6.

This, together with the preceeding result give

Theorem 6.11 (Criterion for Invertibility)
The $n\times n$ matrix A is invertible iff $\det(A) \neq 0$.

Example in class

We now combine some of the above results in a convenient way.

For $1 \leq i \leq n$, let e_i be the vector $(0, 0, \ldots, 0, 1, 0, \ldots, 0)$ with 1 in the i^{th} place.

Remarks 6.12 (Columns, and Expansion by Minors)
1. Since[†] $\det(A) = \det(A^T)$, we can replace all the results about row operations with the corresponding results about column operations.

2. Using the original formula $\det(A) = \Sigma_\sigma \Pi_\sigma(A)\text{sgn}(\sigma)$, we can break up the sum by using those permutations with $\sigma(1) = 1$, then those for which $\sigma(1) = 2$, etc. This gives $\det(A)$ as a sum of n terms, each of which has the form

$$(-1)^{i+1}\det(M_{1i}),$$

where M_{1i} is obtained from A by deleting row 1 and column i. The reason for the sign changes needs a little explanation: The first sign is +1 because deleting 1 from $\{1, 2, \ldots, n\}$ and then identifying $\{2, 3, \ldots, n\}$ with $\{1, 2, \ldots, n-1\}$ causes no sign changes in this part of the expansion, since no inversions are created or destroyed. ($\sigma(1) = 1$ is not an inversion, and the identification above does no harm.) as for the others, $\sigma(1) = i \neq 1$ is an inversion. To see what is happening here, note that $\det(M_{1i})$ is what we would obtain from the *first* summand $\det(N_{11})$ in the matrix B obtained from A by cyclically permuting the rows using the cycle $\tau = (i, i-1, \ldots, 2, 1)$, which has parity $(-1)^{i+1}$. But switching columns using a product of cycles effects *every* summand $\Pi_\alpha(A)$ with the sign change. In other words,

$$\begin{aligned} i\text{ th summand} &= \text{sgn}(\tau) \times \text{ first summand in } \det(B) \\ &= \text{sgn}(\tau)\det(N_{11}) \\ &= \text{sgn}(\tau)\det(M_{ii}) = (-1)^{i+1}\det(M_{1i}) \end{aligned}$$

3. It now follows that we can expand by minors using any row or column, replacing "1" above by "j".

Theorem 6.13 (Characterization of the Determinant)
The determinant function is the one *and only one* real valued function on square matrices that satisfies the following three properties:
(a) multilinearily

$$\det[r_1, \ldots, \lambda r_i + \mu r_j, \ldots, r_n] = \lambda\det[r_1, \ldots, r_i, \ldots, r_n] + \mu\det[r_1, \ldots, r_j, \ldots, r_n]$$

(b) skew-symmetry or **anti-symmetry**

If $i \neq j$, then $\det[r_1, \ldots, r_i, \ldots, r_j, \ldots, r_n] = -\det[r_1, \ldots, r_j, \ldots, r_i, \ldots, r_n]$.

(c) normalcy

$\det[e_1, \ldots, e_n] = 1$.

Proof First, we check that it satisfies these properties. But (a) is Lemma 6.8(b) combined with Proposition 6.9(a), while (b) is Proposition 6.9(b). Finally (c) simply asserts that det(I) = 1.

Next, we show uniqueness. This is essentially saying that the above properties suffice to allow one to compute det of any matrix, which we already know. More precisely, if ϕ was another

[†] See Exercise Set 7.

function which enjoyed these properties, we must show that $\phi(A) = \det(A)$ for every square matrix A.

Thus let A be any square matrix, and let $e_1, e_2, \dots, e_m$ be any sequence of row operations which reduce A. Denote the resulting reduced matrix by B. If $B \neq I$, then properties (a) through (c), (Which hold for both det and ϕ), imply that $\phi(A) = \det(A) = 0$. (See the above Remarks 6.11.) Otherwise, $\det(A)$ is ± 1 divided by the product of all the constants by which we multiplied rows with the e_i, and where the sign is determined by the number of e_i's which were of the row-switching type. Noting that this follows from the properties (a) through (c), the same must be true for $\phi(A)$. Thus again, $\phi(A) = \det(A)$, and we are done. ❾

In view of the theorem, we refer to det as **the unique multilinear antisymmetric normal form** on the set of matrices.

Exercise Set 6
Anton §2.2 #3, 5, 7, 9, 11 (calculate the deternimants by keeping track of the row operations)
Hand In (Value = 15 points)
1. By considering the three types of elementary row operations, show that, if A is any square matrix, and E is any elementary square matrix, then

$$\det(EA) = \det(E)\det(A).$$

2. Show, *without* making any assumptions such as $\det(AB) = \det(A)\det(B)$, that, if A is invertible, $\det(A^{-1}) = 1/\det(A)$.
3. Show that skew-symmetric $n \times n$ matrices have zero determinant if n is odd.

☆ **Considerable Extra Credit** If, without the help of others in the class you can prove—and verbally defend—the following:

1. S'pose $\sigma = (\sigma(1), \sigma(2), \dots, \sigma(n))$. Let $\rho = \sigma$, but with two entries (the i^{th} and the j^{th}, say) switched. Show that $\text{sgn}(\sigma) = -\text{sgn}(\rho)$. [Hint: Determine which pairs $k < l$ get effected if you switch i and j.]

2. Deduce that $\text{sgn}(\sigma \circ \mu) = \text{sgn}(\sigma)\text{sgn}(\mu)$ for every pair of permutations σ, μ.

five points will be added to your final examination. Deadline: One week after the assignment is due.

---

7. Multiplicative Property of the Determinant

We now show another important property of the determinant function, but first we need a little lemma.

Lemma 7.1 Let $A, B \in M(n,n)$. Then, if B fails to be invertible, so do AB and BA.

Proof. We first show that AB is not invertible. S'pose to the contrary that AB *was* invertible. Then, by Proposition 5.4, so is $(AB)^t = B^tA^t$. Let P be its inverse, so $B^tA^tP = I$. But, since B is not invertible, nor is B^t, and so there is a product of elementary matrices (and hence an invertible matrix) S such that $SB^t = I^-$, where I^- has a row of zeros. Now $SB^tA^tP = SI = S$, an invertible matrix. But the left hand side is $(SB^t)A^tP = I^-(A^tP)$, which has a row of zeros. Thus the invertible matrix S has a row of zeros — impossible. Thus AB cannot have been invertible. The proof that BA is not invertible is similar and easier, and left as an exercise (see Exercise Set 7). ❈

Note that one has, *mutatis mutandis*, A not invertible implies AB and BA are not invertible.

Theorem 7.2 Multiplicativity of the Determinant
One has, for any pair A, B of $n\times n$ matrices,

$$\det(AB) = \det(A)\det(B).$$

Proof We consider two cases.

Case 1. A not invertible

In this case, $\det(A) = 0$, (Theorem 6.11). Further, the above lemma implies that AB is not invertible either. Thus $\det(AB) = 0$ too. Thus $\det(AB)$ and $\det(A)\det(B)$ are equal, being both zero.

Case 2. A invertible

In this case, A is a product, $E_1E_2. \ . \ . E_m$ of elementary matrices. By repeated use of Exercise 1 in the above problem set, one now gets

$$\begin{aligned}\det(AB) &= \det(E_1E_2 \ . \ . \ . \ E_mB)\\ &= \det(E_1)\det(E_2 \ . \ . \ . \ E_mB) \text{ (applying it once)}\\ &= \det(E_1)\det(E_2)\det(E_3. \ . \ . E_mB) \text{ (applying it again)}\\ &= \ . \ . \ . \ .\\ &= \det(E_1)\det(E_2). \ . \ .\det(E_m)\det(B).\end{aligned}$$

But, by the same argument,

$$\begin{aligned}\det(A) &= \det(E_1E_2 \ . \ . \ . \ E_m)\\ &= \det(E_1)\det(E_2 \ . \ . \ . \ E_m)\\ &= \ . \ . \ . \ .\\ &= \det(E_1)\det(E_2). \ . \ .\det(E_m).\end{aligned}$$

Combining these gives

$$\det(AB) = \det(E_1)\det(E_2). \ . \ .\det(E_m) \det(B) = \det(A) \det(B),$$

as required. [a]

Corollary 7.3 (A New Criterion for Invertibility)
The square matrix A is invertible iff $\det(A) \neq 0$. If this is the case, then $\det(A^{-1}) = (\det(A))^{-1}$. ✄

Proof in class (Note that we have already established this result without using the theorem.)

Exercise Set 7
Complete the proof of Lemma 7.1.

Hand In (Value = 10)
1. (a) Show that the transpose of an elementary matrix is also an elementary matrix.
(b) Prove that $\det(A^T) = \det(A)$ for eny square matrix A.
(c) Deduce that, in the statement of Theorem 5.8, the word ”row” can be replaced throughout by the word ”column”.

---

8. Vector Spaces

Definitions 8.1 We take $\mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}\}$, the set of all n-tuples of real numbers, and call it **Euclidean n-space**. We refer to an element $(x_1, x_2, \ldots, x_n)$ of $\mathbb{R}^n$ as a **vector**. If $\boldsymbol{v} = (x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, then we refer to the x_i ($i = 1, \ldots, n$) as the **coordinates** of **v.**

For example, Euclidean 1-space is the ”real line”, Euclidean 2-space is the ”xy-plane”, and the vectors are just the points themselves, Euclidean 3-space is what you have previously called ”3-space”, (the set of triples of real numbers).

Definitions 8.2 Let $\boldsymbol{v} = (v_1, v_2, \ldots, v_n)$ and $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ be vectors in $\mathbb{R}^n$ and $\lambda \in \mathbb{R}$. Then define their **sum**, $\boldsymbol{v} + \boldsymbol{w}$, to be given by

$$\boldsymbol{v} + \boldsymbol{w} = (v_1+w_1, v_2+w_2, \ldots, v_n+w_n)$$

and define the **scalar multiple**, $\lambda\boldsymbol{v}$ to be given by

$$\lambda\boldsymbol{v} = (\lambda v_1, \lambda v_2, \ldots, \lambda v_n).$$

Finally, by $-\boldsymbol{v}$, we mean $(-1)\boldsymbol{v}$, and by $\boldsymbol{v} - \boldsymbol{w}$, we mean $\boldsymbol{v} + (-\boldsymbol{w})$.
(Thus we just add coordinates to get the sum, and we multiply through by λ to get the scalar multiple.)

Denote the **zero vector** (0,0, ... ,0) by **0**.

Example 8.3 In $\mathbb{R}^4$, one has
$(2, -1, 0, 9) + (1, 2, 5, \pi) = (3, 1, 5, 9+\pi)$, and
$11(1, 0, 0, 9) - (4, 1, 2, 0) = (7, -1, -1, 99)$,
$(1, 2, 3, 4) - (3, 4, 5, 6) = (-2, -2, -2, -2)$.

Note that the above coincides with the usual addition of vectors in 3-space and 2-space, and with plain old addition of numbers in 1-space.

Examples of geometric interpretation of addition & scalar multiplication—in class.

Proposition 8.4 (Properties of Vectors in $\mathbb{R}^n$)
Let $\boldsymbol{u}$, $\boldsymbol{v}$ and $\boldsymbol{w} \in \mathbb{R}^n$, and let $\lambda, \mu \in \mathbb{R}$ Then:

(a) $\boldsymbol{u} + \boldsymbol{v} = \boldsymbol{v} + \boldsymbol{u}$;	(commutativity of +)
(b) $\boldsymbol{u} + (\boldsymbol{v} + \boldsymbol{w}) = (\boldsymbol{u} + \boldsymbol{v}) + \boldsymbol{w}$	(associativity of +)
(c) $\boldsymbol{u} + \mathbf{0} = \mathbf{0} + \boldsymbol{u} = \boldsymbol{u}$;	(additive identity)
(d) $\boldsymbol{u} + (-\boldsymbol{u}) = (-\boldsymbol{u}) + \boldsymbol{u} = \mathbf{0}$;	(additive inversion)
(e) $\lambda(\boldsymbol{u} + \boldsymbol{v}) = \lambda\boldsymbol{u} + \lambda\boldsymbol{v}$;	(distributivity of scalar mult.)
(f) $\lambda(\mu(\boldsymbol{u})) = (\lambda.\mu)\boldsymbol{u}$;	(associativity of scalar mult.)
(g) $(\lambda + \mu)\boldsymbol{v} = \lambda\boldsymbol{v} + \mu\boldsymbol{v}$;	(distr. over scalar addition)
(h) $1.\boldsymbol{v} = \boldsymbol{v}$	(unitality of scalar mult.)
(i) $0.\boldsymbol{v} = \mathbf{0}$.	(annihilation by 0).

Proof. This is just a special case of Proposition 1.5, as we may think of vectors in $\mathbb{R}^n$ as $1\times n$ matrices. ❇

These properties are actually very important, and we generalize as follows.

Definition 8.5 A **vector space over the reals** is a set V, together with a specified rule for addition and multiplication by elements in $\mathbb{R}$. Addition and scalar multiplication are, in addition, required to satisfy the following rules: If u, v, and $w \in V$ and λ, $\mu \in \mathbb{R}$, then one requires that:

(a) $u + v = v + u$; (commutativity of +)

(b) $u + (v + w) = (u + v) + w$ (associativity of +)

⟶ **(c)** There exists an element $0 \in$ V such that $u + 0 = 0 + u = u$ for all $u \in V$; (we call 0 the additive identity)

⟶ **(d)** For each $u \in V$, there exists a corresponding element $-u \in V$ with the property that $u + (-u) = (-u) + u = 0$; (we call -u the additive inverse of u)

(e) $\lambda(u + v) = \lambda u + \lambda v$; (distributivity of scalar mult.)

(f) $\lambda(\mu(u)) = (\lambda.\mu)v$; (associativity of scalar mult.)

(g) $(\lambda + \mu)v = \lambda v + \mu v$; (distr. over scalar addition)

(h) $1.v = v$ (unitality of scalar mult.)

(i) $0.v = 0$. (annihilation by 0).

Note Since vector addition is an operation on the vector space V, it is required that the sum $v+w$ of two elements v and w of V is also an element of V. Similarly for scalar multiplication. this property is called **closure.**

Closure:	
If v and $w \in V$, then $v+w \in V$	(Closure under addition)
If $v \in V$ and $\lambda \in \mathbb{R}$, then$\lambda v \in V$	(Closure under scalar multiplication)

Examples 8.6

(a) $\mathbb{R}^n$, with the addition and scalar multiplication specified above, enjoys these properties by Proposition 8.4, so it is a vector space.

(b) The set M(m,n) of $m\times n$ matrices, with addition and multiplication specified in Definition 1.3, also enjoys these properties by Proposition 1.5, so it too is a vector space.

(c) Let $C(a, b)$ denote the set of continuous real-valued functions on the interval $(a,b) \subset \mathbb{R}$. Thus for a function f to be in $C(a, b)$, one requires that it be defined on (a, b) and that it be continuous at each point in (a, b). If f and g are in $C(a, b)$, define $f+g \in C(a, b)$ by the rule

$$(f+g)(x) = f(x) + g(x),$$

and if $\lambda \in \mathbb{R}$ define a function λf by the rule

$$(\lambda f)(x) = \lambda.f(x).$$

Then one has the zero function $0 \in C(a, b)$ given by $0(x) = 0$, and also, for each $f \in C(a, b)$, the additive inverse $-f \in C(a,b)$ given by

$$(-f)(x) = -f(x).$$

That $C(a, b)$ with this structure is indeed a vector space is verified in Problem Set 10.

(d) Let $C^1(a, b)$ denote the set of differentiable real-valued functions on (a, b), so that we require that the first derivative of f exist everywhere on (a, b) for each $f \in C^1(a, b)$. Addition and scalar multiplication are given as in (c). Similarly, let $C^n(a, b)$ denote the set of *n times differentiable* real-valued functions on (a, b), so that we require that the nth derivative exist everywhere in (a, b). Finally, guess what $C^\infty(a, b)$ is.

(e) Let $\mathbb{R}[x]$ denote the set of all **polynomials with real coefficients** in the "indeterminate" x. Thus elements of $\mathbb{R}[x]$ are *expressions* of the form

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$$

with each $a_i \in \mathbb{R}$ and all but finitely many of the a_i zero.[†] (Thus we think of the polynomial as "stopping" somewhere. (eg. $4x^2 + 3x - 100$, $x^3 - x$, etc.) We add polynomials by adding corresponding coefficients, so

$$(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots) + (b_0 + b_1x + b_2x^2 + \ldots + b_nx^n + \ldots) =$$
$$(a_0+b_0) + (a_1+b_1)x + (a_2+b_2)x^2 + \ldots + (a_n+b_n)x^n + \ldots$$

(eg. $(x^2 + x) + (3x^3 - 2x^2 - 4x + 8) = 3x^3 - x^2 - 3x + 8.$)

For scalar mult, we define

$$\neg(a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots) = -a_0 - a_1x - a_2x^2 - \ldots - a_nx^n - \ldots\ldots$$

With these rules, $\mathbb{R}[x]$ is a vector space (this exercise set #2).

(f) Define the set of **power series with real coefficients**, which we denote by $\mathbb{R}[[x]]$, exactly as in (e), but with the requirement that all but finitely many of the a_i be zero dropped. For example, an element of $\mathbb{R}[[x]]$ is

$$1 + x + x^2 + \ldots + x^n + \ldots.$$

Addition and scalar multiplication are then defined exactly as for polynomials.

(g) Let P be any plane through the origin in $\mathbb{R}^3$, so suppose P is the set of all points (x, y, z) such that $ax + by + cz = 0$ for some given constants $a, b, c \in \mathbb{R}$. Then P is a vector space; We add points in P by the rule

$$(x, y, z) + (x', y', z') = (x+x', y+y', z+z'),$$

and scalar mult is given by

$$\lambda(x, y, z) = (\lambda x, \lambda y, \lambda z).$$

Note that if (x,y,z) and (x',y',z') are points in P, then so is $(x+x',y+y',z+z')$, since the new point has

$$a(x+x') + b(y+y') + c(z+z') = ax + by + cz + ax' + by' + cz' = 0 + 0 = 0.$$

Similarly for scalar mult. That the axioms continue to be obeyed follows from the fact that addition and scalar mult. are inherited from the corresponding operations in $\mathbb{R}^3$, where they do indeed satisfy the axioms.
(illustration in class)

[†] This is analogous to the definition of a complex number as an "expression" of the form $ai+b$. Kolman bungles it on page 100 where he defines a polynomial as a function. It isn't.

(h) Similarly, a *line* through the origin in $\mathbb{R}^3$ consists of all points (x, y, z) satisfying the *system* of equations

$$ax + by + cz = 0$$
$$dx + ey + fz = 0$$

for some given a, b, c, d, e, f such that (d, e, f) is not a scalar multiple of (a, b ,c) in $\mathbb{R}^3$. (Why?)
Roughly speaking, a vector space is anything in which addition and scalar multiplication make sense and behave ”properly”.

(i) The set Q of all points in the first quadrant of $\mathbb{R}^2$ is not a vector space, since the (usual) addition and scalar multiplication don't work well there. Eg., $-1.(1.1) = (-1,-1)$, but this is not an element of Q.

(j). There is a silly little vector space, called $\{0\}$, which contains a single element called 0, and which has the addition and scalar mult. rules:

$$0 + 0 = 0$$
$$\lambda.0 = 0.$$

It can (easily!) be checked that the rules of a vector space are satisfied by $\{0\}$. We call $\{0\}$ the **trivial vector space**, and picture it as a single point *. (Actually, this vector space plays an important theoretical role, as we shall see below.) For example, the origin in $\mathbb{R}^3$ is a trivial vector space, consisting of the single element (0, 0, 0).

Exercise Set 8
Anton §5.1 #1–15 odd

Hand In (Value = 20)
1. Verify that $C(a, b)$ is a vector space (cf. Example 8.6(c)).
2. Verify that $\mathbb{R}[x]$ is a vector space (cf. Example 8.6(e)).
3. Give a geometric reason answering the question in Example 8.6(h).
4. **(a)** Show that axiom 8.5(i) can be deduced from the other axioms.
(b) Use the axioms of a vector space to show that if u is an element of the vector space V, then $(-1)u = -u$.
(c) Show that the additive inverse of every element in a vector space is unique.

---

9. Subspaces and Spans

Definition 9.1 A subset $W \subset V$ of the vector space V is called a **subspace of** V if the set of elements in W form a vector space in their own right under the operations in V.

In order to check whether a given subset of V is a subspace of V, we use the folowing criterion.

Proposition 9.2 (Test for a Subspace)
A subset W of the vector space V is a **subspace** iff the following conditions hold.
 (a) If u and v are in W, then so is $u + v$;
 (b) If $u \in W$, and $\lambda \in R$ then $\lambda u \in W$.
(That is, W is a subset which is **closed under addition and scalar multiplication.**)

Proof S'pose $W \subset V$ is a vector subspace. Then (a) and (b) must hold, since W, being a vector space, contains the sums and scalar multiples of its elements.

Conversely, s'pose $W \subset V$ satisfies (a) and (b). Then W contains the negatives of its elements, since, for $w \in W$, $-w = (-1).w \in W$ by (b). Thus W contains 0 too, since $0 = w + (-w) \in W$ by (a). Now all the remaining axioms hold in W since they hold in V. ➢

Examples 9.3
(a) Let P be any plane through the origin in $\mathbb{R}^3$. Then P is a subspace of $\mathbb{R}^3$ since: If u and $v \in P$, then so are $u+v$ and λu for any λ, as we verified in Example 8.6(g).

(b) A polynomial of degree $\leq n$ over $\mathbb{R}$ is a polynomial

$$a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$$

over $\mathbb{R}$ with $a_{n+1} = a_{n+2} = \ldots = 0$, (so we can stop at the x^n term). Claim: the set $\mathbb{R}_n[x]$ of polynomials of degree $\leq n$ over $\mathbb{R}$ is a vector subspace of $\mathbb{R}[x]$. We check this in class.

(c) The set $D(n)$ consisting of all $n\times n$ diagonal matrices is a subspace of the vector space $M(n,n)$ of all $n\times n$ matrices. (This exercise set).

(d) Let A be an $m\times n$ matrix and let $AX = 0$ be any homogeneous system of equations (in matrix form). Consider the set W of all solutions

$$(s_1, s_2, \ldots, s_n),$$

(previously known as the solution set). Claim: This is in fact a subspace of $\mathbb{R}^n$.
Proof We need only show that the sum of two solutions is again a solution, and that scalar multiples of solutions are solutions. But this is precisely what Lemma 3.2 said ➷

In view of this we henceforth refer to the solution set as the **solution space**; it's been promoted.

Definition 9.4 Let $\mathscr{S} = \{v_1, v_2, \ldots, v_n, \ldots\}$ be a set of any n vectors in the vector space V. Then the **span of $\mathscr{S}$** is the set $\langle\mathscr{S}\rangle$ consisting of all finite linear combinations of the v_i. That is,

$$\langle\mathscr{S}\rangle = \{\alpha_1v_1 + \alpha_2v_2 + \ldots + \alpha_nv_n \mid a_i \in \mathbb{R},\ n \geq 1\}.$$

We sometimes write $\langle\mathscr{S}\rangle$ as $\langle v_1, v_2, \ldots, v_n, \ldots\rangle$.

Examples in class: Illustration of spans in $\mathbb{R}^3$.

Examples 9.5
(a) Let $V = \mathbb{R}^3$ and let $\mathscr{S} = \{(1,0,0), (1,1,0)\}$. Then the span of $\mathscr{S}$ is $\{(a, b, 0) : a, b \in \mathbb{R}\}$.

Picture in class

Proof $\langle \mathscr{S} \rangle = \{\alpha(1, 0, 0) + \beta(1, 1, 0) : a, b \in \mathbb{R}\}$ by definition of the span
$= \{(a, b, 0) : a, b \in \mathbb{R}\}$ by definition of addition of vectors.

(b) Let $V = \mathbb{R}^3$ and let $\mathscr{S} = \{(1, 0, 0), (1, 1, 0), (0, 0, 1)\}$. Then the span of $\mathscr{S}$ is the whole of $\mathbb{R}^3$.

(c) Let $V = \mathbb{R}[x]$, and let $\mathscr{S} = \{1, x, x^2, \dots, x^n\}$. Then $\langle \mathscr{S} \rangle$ is the subspace $\mathbb{R}_n[x]$.
Proof Exercise Set 9.

Proposition 9.6 (Spans are Subspaces)
If $\mathscr{S} = \{v_1, v_2, \dots, v_n\}$ is any set of n vectors in V, then $\langle \mathscr{S} \rangle$ is a subspace of V.

Proof We need only show closure under (+) and (.). Thus let u and $w \in \langle \mathscr{S} \rangle$. This means that

$$u = \alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$$

and

$$w = \beta_1 v_1 + \beta_2 v_2 + \dots + \beta_n v_n$$

for some α_i and β_i in $\mathbb{R}$. Adding gives

$$u + w = (\alpha_1+\beta_1)v_1 + (\alpha_2+\beta_2)v_2 + \dots + (\alpha_n+\beta_n)v_n,$$

which is again in $\langle \mathscr{S} \rangle$. Similarly for scalar multiples. ➡

In view of this, we also refer to $\langle \mathscr{S} \rangle$ as the **subspace spanned by the set $\mathscr{S}$.**

It may happen that $\langle \mathscr{S} \rangle$ is the whole of V, as was seen above. Then we refer to $\mathscr{S}$ as a **spanning set of V**.

Proposition 9.7 (Row Operations Don't Effect the Span)
If A and B are two row-equivalent $m\times n$ matrices, then the span of the rows of A = the span of the rows of B.

Proof If e is an elementary row operation, then each row in $e(A)$ is a linear combination of the rows in A. (Check each one to see why.) Thus, anything in the span of the rows of $e(A)$ is also in the span of the rows of A, since it is a linear combination of linear combinations . . .
Conversely, since $A = e^{-1}(e(A))$, where e^{-1} is also an elementary row operation, everything in the span of the rows of A is also in the span of the rows of $e(A)$. Thus, the spans of the rows of A and $e(A)$ are identical.

Thus, doing an elementary row operation doesn't alter the span of the rows. Thus, doing a sequence of row operations doesn't alter the span of the rows. ✥

It follows that, to get a good idea of what the span of a bunch of vectors looks like, put 'em in a matrix and row reduce.

Examples of spans which span $\mathbb{R}^3$ and a proper subspace in class

We now look at how to get new subspaces from old.

Definition 9.8 Let U and W be subspaces of the vector space V. Then the **sum** of U and W is given by

$$U+W = \{u+w \mid u \in U, w \in W\}.$$

Examples in class; the sum of two lines; the sum of a line and a plane.

Proposition 9.9 (Sums of Subspaces)
The sum of any two subspaces of V is a subspace of V. Moreoever, it is the smallest subspace of V that includes U and W. That is, if X is a subspace of V that includes U and W, then $X \supset U+W$.

Exercise Set 9
Anton §5.2, #1, 3, 5

Hand In: (Value = 25)
1. Verify that $D(n) \subset M(n,n)$ is a subspace (9.3(c)).
2. Verify that $C^1(a,b) \subset C(a,b)$ is a subspace, quoting some theorems from Calculus if necessary.
3. Verify the claim in Example 9.5 (c).
4. **(a)** Prove that the intersection of two subspaces of V is a subspace of V.
(b) If W_1 and W_2 are subspaces of V, is the union $W_1 \cup W_2$ necessarily a subspace of V? Prove, or give a counterexample.
5. Two subspaces U and W are **complementary subspaces of V** if two conditions hold:
(1) $U+W = V$
(2) $U \cap W = \{0\}$.
Find a linear complement to the solution space of $2x - 3y + 4z = 0$ in $\mathbb{R}^3$.

---

10. Linear Independence

Consider the following three vectors in $\mathbb{R}^3$:

$v_1 = (1, -1, 0)$, $v_2 = (0, 1, 2)$ and $v_3 = (1, -3, -4)$.

It just so happens that $v_3 = v_1 - 2v_3$. (If you don't believe this, then check for yourself!) In other words, v_3 is a linear combination of the other two, (i.e.. is in the span of the other two) since we can write

$v_3 = \lambda_1 v_1 + \lambda_2 v_3$,

where here, $\lambda_1 = 1$ and $\lambda_2 = -2$. Thus in a manner of speaking, the third vector "depends" on the other two. Incidentally, we can also solve for v_2 and write it as a linear combination of the others. On the other hand, now consider the three vectors

$w_1 = (1, -1, 0)$, $w_2 = (0, 1, 2)$ and $w_3 = (1, -3, 0)$.

No matter how hard we try, we cannot express one of these as a linear combination of the other two. (We'll see how to tell this at a glance later.) We therefore make the following definition.

Preliminary Definition 10.1 S'pose V is a vector space, and $v_1, v_2, \ldots, v_r$ are in V. If one of them, is in the span of the others, then we say that the vectors $v_1, v_2, \ldots, v_r$ are **linearly dependent**. Otherwise, we say that they are **linearly independent**.

Thus, for example, the vectors v_1, v_2, v_3 above are linearly dependent, whereas w_1, w_2, w_3 are linearly independent. It turns out that this definition of linear dependence is a little cumbersome in practice (which is why we called it "preliminary") so we develop another.

S'pose that the collection of vectors $v_1, v_2, \ldots, v_r$ is linearly dependent, so that one of them— let us say v_1— can be written as a linear combination of the others. Thus:

$v_1 = \lambda_2 v_2 + \lambda_3 v_3 + \ldots + \lambda_r v_r$

for some scalars λ_i. Rewriting gives

$-v_1 + \lambda_2 v_2 + \lambda_3 v_3 + \ldots + \lambda_r v_r = 0$.

This implies that the equation

$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \ldots + \mu_r v_r = 0 \quad \ldots\ldots (*)$

has at least one non-zero solution for the μ_i, namely, $(-1, \lambda_2, \lambda_3, \ldots \lambda_r)$. Conversely, if equation (*) possesses a non-zero solution, say

$(\rho_1, \rho_2, \ldots, \rho_r)$

with, say, $\rho_i \neq 0$, then we can solve equation (*) for v_i , thus getting vi as a linear combination of the others.

Thus: linear dependence amounts to saying that the equation

$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \ldots + \mu_r v_r = 0 \quad \ldots\ldots (*)$

has at least one non-zero solution for the μ_i

When this can't be done, i.e., when the v_i's are independent, then equation (*) has only the zero solution. Thus we make the following fancy definition.

Fancy Definition 10.2 Let V be a vector space, and let $v_1, v_2, \ldots, v_r$ be in V. Then they are said to be **linearly independent** if the equation

$$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \ldots + \mu_r v_r = 0$$

has only the zero solution for the μ_i . Otherwise, they are said to be **linearly dependent.**

This gives us the following:

Test for Linear Independence

1. If the vectors you are looking at happen to be in $\mathbb{R}^n$, then use Stef's sure-fire method: put them in a matrix and row reduce (see below).

2. To show that a given collection $v_1, v_2, \ldots, v_r$ of vectors is linearly independent, set

$$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \ldots + \mu_r v_r = 0$$

and prove that each of the μ_i must be 0.

3. To show that a given collection $v_1, v_2, \ldots, v_r$ of vectors is linearly dependent, find scalars

$$\mu_1, \mu_2, \ldots, \mu_r,$$

not all zero, such that

$$\mu_1 v_1 + \mu_2 v_2 + \mu_3 v_3 + \ldots + \mu_r v_r = 0.$$

Examples 10.3

(a) The vectors **i** = (1, 0, 0), **j** = (0, 1, 0) and **k** = (0, 0, 1) are linearly independent in $\mathbb{R}^3$. Indeed, if

$$\mu_1(1, 0, 0) + \mu_2(0, 1, 0) + \mu_3(0, 0, 1) = (0, 0, 0) \longleftarrow \text{the zero vector,}$$

then multiplying through by the scalars μi and adding gives

$$(\mu_1, \mu_2, \mu_3) = (0, 0, 0)$$

so

$$\mu_1 = \mu_2 = \mu_3 = 0,$$

showing that equation (*) has only the zero solution.

(b) The vectors $v_1 = (2, -1, 0, 3)$, $v_2 = (1, 2, 5, -1)$ and $v_3 = (7, -1, 5, 8)$ are linearly dependent in $\mathbb{R}^4$, since $3v_1 + v_2 - v_3 = 0$. (Check !)

(c) Important Example: If B is any matrix in reduced row-echelon form, then the non-zero rows of B are independent. Reason: The leading entry of each of these vectors is in a different slot, and the other vectors have zeros in that slot. So, if a linear combination of them is zero, then the coefficients must be zero.

(d) Discussion of Wronskian in class

In order to tell easily whether a bunch of vectors in $\mathbb{R}^n$ is linearly independent, we use the following lemma.

Proposition 10.4 (Stef's Sure Fire Method for Checking Independence)

The collection $v_1, v_2, \ldots, v_r$ of vectors in $\mathbb{R}^n$ is linearly independent iff the matrix A whose rows are the v_i reduces to a matrix B with no rows of zeros.

Proof If we get a row of zeros when we reduce A, then this means that one row was a combination of the others, so that the rows of A were linearly dependent. Conversely, if the rows of A are linearly dependent, then there is a non-trivial linear combination of the rows that results in zero, and so there is a sequence of row operations that results in a row of zeros. This finishes the proof, if we accept that, no matter how you reduce the matrix, you will always get a row of zeros. In fact, this is true because the row-reduced form of a matrix is unique. However, we have not proved this fact. Thus, the proof rests on the following claim:[‡]

Claim: If reducing a matrix in a certain way leads to a row of zeros, then no matter how you reduce the matrix, you must always get a row of zeros.

Proof of Claim: If you reduce it and obtain rows $v_1, v_2, \ldots, v_n$, with $v_n = 0$, then

$$V = \langle v_1, \ldots, v_{n-1} \rangle$$

is the span of the original rows. If we now find a Martian named Zort to reduce it, and Zort gets a matrix with rows $w_1, w_2, \ldots, w_n$ (no rows of zeros) then

$$V = \langle v_1, v_2, \ldots, v_{n-1} \rangle = \langle w_1, w_2, \ldots, w_n \rangle.$$

I claim that the w_i must be linearly dependent, contradicting Example 10.3 part (c) above. To show this, we appeal to the following result, which is Lemma 11.4 below:

If $\{v_1, v_2, \ldots, v_r\}$ spans the vector space V, then any collection of more than r vectors in V is linearly dependent. ❄

Corollary 10.5 In $\mathbb{R}^n$, any set of more than n vectors is linearly dependent. ✙

Question: If we do reduction without permitting rows to get switched, then do the row(s) of zeros in B correspond to vectors that are linear combinations of the others?
Answer: Not necessarily: a row swap can be done by a sequence of other row operations (how?).
Question: So how can we find out which rows were linear combinations of the others?
Answer: Instead of doing reduced row-echelon form, we just clear *below* the leading entries. Then, if you get a row of zeros, that row must have been a linear combination of the rows above. Also, completing the row reduction to reduced form cannot introduce any additional rows of zeros, since clearing above leading entries won't wipe out any other leading entries.

[‡] which was proved earlier for *square* matrices (see Exercise set 4 #4).

Exercise Set 10
Anton §5.3 #1–15 odd

Hand In (Value = 15)

1. If V is a vector space, and $S_1 \subset S_2 \subset V$, show that:

(a) If S_1 is linearly dependent, so is S_2

(b) If S_2 is linearly independent, so is S_1 ("subsets of independent sets are independent").

2. Let $v_1, v_2, \ldots, v_n, \ldots$ be a sequence of vectors in V such that no vector in the sequence is a linear combination of its predecessors. Show that $\{v_1, v_2, \ldots, v_n, \ldots\}$ is a linearly independent set.

3. Let $\mathscr{S}$ be a collection of linearly independent vectors, and let v be a vector not in the span of S. Prove that $S \cup \{v\}$ is linearly independent.

11. Bases

Definition 11.1 A **basis** for the vector space V is a linearly independent set of vectors whose span is (the whole of) V.

To Show That a Set $\mathscr{B}$ is a Basis for V
1. Show that $\mathscr{B}$ is linearly independent—see the previous section for methods.
2. Show that $\mathscr{B}$ spans V; that is, let $v \in V$ be arbitrary, and show that v can be written as a linear combination of vectors in $\mathscr{B}$.

Examples 11.2
(a) $\mathbb{R}^2$ has a basis $\mathscr{B} = \{\boldsymbol{i}, \boldsymbol{j}\}$, where $\boldsymbol{i} = (1, 0)$ and $\boldsymbol{j} = (0, 1)$.
(b) $\mathbb{R}^3$ has a basis $\mathscr{B} = \{\boldsymbol{i}, \boldsymbol{j}, \boldsymbol{k}\}$, where $\boldsymbol{i} = (1, 0, 0)$, $\boldsymbol{j} = (0, 1, 0)$ and $\boldsymbol{k} = (0, 0, 1)$.
(c) $\mathbb{R}^1 = \mathbb{R}$ has a basis $\mathscr{B} = \{\boldsymbol{i}\}$, where $\boldsymbol{i} = (1)$.
(d) In general, $\mathbb{R}^n$ has a basis $\mathscr{B} = \{e_1, e_2, \ldots, e_n\}$, where

$$e_i = (0, 0, \ldots, \underset{\substack{\uparrow \\ ith\ place.}}{1}, \ldots, 0)$$

Proof in homework.
(e) $\mathscr{B} = \{1, x, x^2, \ldots, x^n\}$ is a basis for $\mathbb{R}_n[x]$.
Indeed, we have already seen that $\mathscr{B}$ is a spanning set. To show independence, if $\mu_0 + \mu_1 x + \ldots + \mu_n x^n = 0$ (the zero *polynomial)*, this means that all the μ_i are zero; for two polynomials in $\mathbb{R}^n[x]$ to be the same, *all* their coefficients must agree.
(f) $\mathscr{B} = \{1, x, x^2, \ldots, x^n, \ldots\}$ is a basis for $\mathbb{R}[x]$. (Exercise Set 11.)

Note on spans: Even if $\mathscr{S}$ is an infinite set, $\langle S\rangle$ by definition consists only of *finite* linear combinations $\lambda_1 v_{r_1} + \lambda_2 v_{r_2} + \ldots + \lambda_s v_{r_s}$.

(g) $\mathscr{B} = \{(1, 0, 9), (-1, 2, 3), (4, 101, 22)\}$ is a basis for $\mathbb{R}^3$.

It seems to suggest itself that the "*dimension*" of V is the number of vectors in a basis. But before we can talk about dimension we need an auxiliary result.

Lemma 11.3 (Sets Bigger than Spanning Sets are Dependent)
If $\{v_1, v_2, \ldots, v_r\}$ spans the vector space V, then any collection of more than r vectors in V is linearly dependent.

Proof
S'pose $\mathscr{W} = \{w_1, w_2, \ldots, w_n\}$ is any collection of vectors in V with $n > r$. To prove that W is dependent, I must produce constants $\mu_1, \mu_2, \ldots, \mu_r$, not all zero, such that

$$\mu_1 w_1 + \mu_2 w_2 + \mu_3 w_3 + \ldots + \mu_n w_n = 0 \qquad (*)$$

Now, the fact that $\{v_1, v_2, \dots, v_r\}$ spans V implies that there are constants a_{ij} such that

$$\begin{aligned}
w_1 &= a_{11}v_1 + a_{12}v_2 + \dots + a_{1r}v_r \\
w_2 &= a_{21}v_1 + a_{22}v_2 + \dots + a_{2r}v_r \\
&\dots \\
w_n &= a_{n1}v_1 + a_{n2}v_2 + \dots + a_{nr}v_r
\end{aligned}$$

Substituting these in (*) gives the big equation

$$\begin{aligned}
&\mu_1(a_{11}v_1 + a_{12}v_2 + \dots + a_{1r}v_r) \\
&+ \mu_2(a_{21}v_1 + a_{22}v_2 + \dots + a_{2r}v_r) \\
&+ \dots \\
&+ \mu_n(a_{n1}v_1 + a_{n2}v_2 + \dots + a_{nr}v_r) = 0
\end{aligned}$$

that we must solve for the μ_i. In fact, we can do better: we can solve the system of r equations in n unknowns obtained from reading the equation down the columns:

$$\begin{aligned}
a_{11}v_1\mu_1 + a_{21}v_1\mu_2 + \dots + a_{n1}v_1\mu_n &= 0 \\
a_{12}v_2\mu_1 + a_{22}v_2\mu_2 + \dots + a_{n2}v_2\mu_n &= 0 \\
&\dots \\
a_{1r}v_r\mu_1 + a_{2r}v_r\mu_2 + \dots + a_{nr}v_r\mu_n &= 0
\end{aligned}$$

and obtain infinitely many non-zero solutions for the μ_i, proving the claim. ❄

Corollary 11.4 (Number of Elements in a Basis)
If V has a finite basis, then any two bases of V have the same number of elements.

Proof. Let $\mathscr{B} = \{v_1, v_2, \dots, v_r\}$ be a finite basis, and let $\mathscr{C} = \{w_1, w_2, \dots, w_n, \dots\}$ be any other basis for V.

Case 1. $\mathscr{C}$ is infinite Let $\mathscr{D} = \{w_1, w_2, \dots, w_{r+1}\}$ consist of any $r+1$ distinct elements of $\mathscr{C}$. Then $\mathscr{D}$ remains linearly independent by Exercise Set 10 #1. Since $\mathscr{B}$ is a basis, it spans V, and thus, by the lemma, the larger set $\mathscr{D}$ cannot be linearly independent, a contradiction.

Case 2. $\mathscr{C}$ is finite

L et $\mathscr{C} = \{w_1, w_2, \dots, w_n\}$. We must show $n = r$. Assume this is not the case. By re-naming the bases if necessary, we can assume $n > r$. Since $\mathscr{B}$ is a basis, it spans V, and thus, by the lemma, the larger set $\mathscr{C}$ cannot be linearly independent, a contradiction. ❐

Definition 11.5 The **dimension** of the vector space V is the number of vectors in any basis $\mathscr{B}$; if a basis $\mathscr{B}$ contains infinitely many vectors, then we say that V is **infinite dimensional**.

By the lemma (see also Exercise Set 11 #2), the notion of dimension is well-defined; that is, a vector space cannot have two different dimensions.

Theorem 11.6 (**Basis Theorem**). The following are equivalent for any collection of vectors $\mathscr{B}$ in the n-dimensional space V.

(a) $\mathscr{B}$ is a basis for V.

(b) $\mathscr{B}$ is a linearly independent spanning set of V.

(c) $\mathscr{B}$ is any collection of n linearly independent vectors in V.

(d) $\mathscr{B}$ is any collection of n vectors which span V.

Proof

(a) ⇒ (b):

This is the definition of a basis.

(b) ⇒ (c):

We are given that $\mathscr{B}$ is linearly independent already, so we need only show that $\mathscr{B}$ has n elements. But we have shown (Corollary 11.5) that if $\mathscr{B}$ is *any* basis of an n-dimensional space (meaning there is *some* basis with n elements), then it has exactly n things in it.

(c) ⇒ (d):

We must show that $\mathscr{B}$ spans V. If $\mathscr{B}$ did not, choose a vector w not in the span of $\mathscr{B}$. Then one can see that $\mathscr{B} \cup \{w\}$ is a collection of $n+1$ independent vectors in the n-dimensional vector space V. (Exercise Set 10 #2) But this contradicts Lemma 11.4.

(d) ⇒ (a):

We must show that any spanning set $\mathscr{B}$ consisting of n vectors is a basis. We already have that it spans V, so we must show that it is l.i. If it *was not* l.i., then one of them is a combination of the others, so throw it away. What is left still spans V (the discarded one can be gotten from the others). If what's left still fails to be l.i., throw another one away. Again, what's left still spans V. Continue going until what's left is l.i. (Eventually, it *must* be, otherwise you eventually get down to one vector, which is automatically l.i.) What is now left is an l.i. set with fewer than n vectors but which still spans V. That is, we now have a basis with $< n$ elements, contradicting 11.4. ❇

Corollary 11.7 (Test for a Basis)

A set $\mathscr{B}$ of n vectors in $\mathbb{R}^n$ *is* a basis iff the matrix A whose rows are the vectors in $\mathscr{B}$ is invertible.

Proof By 5.8, A is invertible iff it is row-equivalent to the identity matrix, which is true iff there are no rows of zeros when we reduce it. By 10.4, this is equivalent to saying that the rows are independent, and hence form a basis by the above theorem. ✱

Note: You can use this to test for a basis in the problems below—just take the determinant or row-reduce.

The following result asserts that if you fix a basis $\mathscr{B}$, then each vector in can be expressed using "coordinates" with respect to that basis.

Proposition 11.8 (Coordinates With Respect to a Basis)

Let $\mathscr{B} = \{v_1, \ldots v_n\}$ be a basis for the vector space V. Then if $v \in V$ is any vector, one can write

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

for some scalars λ_i. Further, the λ_i are unique for a given vector v. We therefore refer to the λ_i as the **coordinates of v with respect to the basis $\mathscr{B}$.**

Proof Since $\mathscr{B}$ is a basis, and hence spans V, it follows that every $v \in V$ is in the span of $\mathscr{B} = \{v_1, \ldots v_n\}$, and hence

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

for some scalars λ_i . To show uniqueness, we suppose that

$$v = \mu_1 v_1 + \mu_2 v_2 + \ldots + \mu_n v_n$$

for some (possibly different) scalars μ_i. Then subtracting, we get

$$0 = (\lambda_1 - \mu_1)v_1 + (\lambda_2 - \mu_2)v_2 + \ldots + (\lambda_n - \mu_n)v_n.$$

Since the v_i are linearly independent, it follows that

$$(\lambda_1 - \mu_1) = (\lambda_2 - \mu_2) = \ldots = (\lambda_n - \mu_n) = 0$$

and hence that $\lambda_i = \mu_i$ for each $i = 1, \ldots, n$, showing uniqueness. ✯

Finally, for those infinite dimensional spoace buffs out there, we have the following:

Definition 11.9 Let **P** be any property, and let S be any set. A subset $A \subset S$ is said to be a **maximal subset** with property **P** if:

(a) A has property **P**;

(b) If $B \subset S$ also has property **P**, and $B \supset A$, then $B = A$.

In words, A has property **P**, and is properly contained in no other subset with property **P**.

A **minimal** subset with property **P** is defined similarly.

For example, by a **maximal independent set** in a vector space V, we mean aa maximal subset of V with the property that it is independent. The following theorem[†] applies to infinite dimensional spaces as well as finite dimensional ones.

Theorem 11.10 (Fancy Basis Theorem). The following are equivalent for any collection of vectors $\mathscr{B}$ in the vector space V.

(a) $\mathscr{B}$ is a basis for V.

(b) $\mathscr{B}$ is a linearly independent spanning set of V.

(c) $\mathscr{B}$ is a maximal linearly independent set.

(d) $\mathscr{B}$ is a minimal spanning set.

Moreover, it can be shown—but the proof is perhaps a little beyond the scope of this course—that:

Theorem 11.11 (Existence of a Basis)
Every vector space has a basis.

[†] not mentioned in Kolman

(Of course, it is a tautology that every *finite dimensional* vector space has a basis...)

Exercise Set 11

Anton §5.4 #1–15 odd

Also[‡] , use Exercise Set 10 #2 to find a basis of the subspace the span of $\{x^2+1,\ x^2+x-1,\ 3x-6,\ x^3+x^2+1,\ x^3\}$ in $\mathbb{R}[x]$.

Hand In (Value = 25)

1. (a) Show that $\mathcal{B} = \{e_1, e_2, \ldots, e_n\}$ in Example 11.2(d) is a basis for $\mathbb{R}^n$.

(b) Example 11.2(f)

2. (a) Show that, if a vector space has an infinite basis, then it has no fninite basis.

(b) Deduce that $\mathbb{R}[x]$ is infinite dimensional.

3. (a) Let W be a subspace of the n-dimensional vector space V with the property that W is also n-dimensional. Prove that $W = V$.

(b) Does part (a) remain true if "n-dimensional" is replaced by "infinite dimensional?" Prove or give a counterexample.

4. Let V be any finite-dimensional vector space. Prove:

(a) Any spanning set of V contains a basis.

(b) Any linearly independent set in V is a contained in a basis.

(**Note** These statements remain true in the infinite dimensional case.The proof of 4(b) in the infinite dimensional case requires Zorn's Lemma. Ask for a discussion...)

5. Let A be an invertible $n \times n$ matrix, and let $\mathcal{B} = \{v_1, \ldots, v_n\}$ be a basis for an r-dimensional subspace of $\mathbb{R}^n$. Show that $\{Av_1, \ldots Av_n\}$ is also a basis for for an r-dimensional subspace of $\mathbb{R}^n$.

[‡] This is from Fraleigh, *Linear Algebra*, 3rd Edition, p. 203.

12. Linear Transformations

If V and W are two vector spaces (possibly even the same vector space), then if f: is a function whose domain is V and whose codomain is W, we say that f is a **mapping from V to W,** and write

$$f\colon V \longrightarrow W.$$

Examples 12.1

(a) Define $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ by $f(x,y) = (x^2, xy, x+y)$. Note here that the *Image* (i.e.. *range*) of f is not the whole of $\mathbb{R}^3$. It doesn't matter. All this means is that f assigns to each member of $\mathbb{R}^2$ a member of $\mathbb{R}^3$.

(b) If V is any vector space, define $\mathbf{1}\colon V \longrightarrow V$ by $\mathbf{1}(v) = v$. The map $\mathbf{1}$ is called the *identity map on V*. (For example $f\colon$ R $\longrightarrow$ R given by $f(x) = x$ is the identity map on the vector space R.)

(c) *An extremely important example*. If A is any $m\times n$ matrix, then A determines an associated map

$$\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

as follows. If $(x_1, x_2, \ldots, x_n) \in \mathbb{R}^n$, define

□ $$\hat{A}(x_1, x_2, \ldots, x_n) = [A\cdot(x_1, x_2, \ldots, x_n)^t]^t$$

where $\cdot$ is matrix multiplication. For example, if A is the 2×2 matrix

$$A = \begin{bmatrix} 1 & 2 \\ 1 & -3 \end{bmatrix},$$

then $\hat{A}(x,y) = (x+2y, x-3y)$.

(d) Define $\varepsilon\colon \mathbb{R}[x] \longrightarrow \mathbb{R}$ by

$$\varepsilon(a_0 + a_1x + \ldots + a_nx\mathrm{n} + \ldots) = a_0 + a_1 + \ldots + a_n + \ldots$$

(Since all but finitely many of the a_i' s are zero, the sum on the right is really a finite sum, so there is no problem with convergence, etc.) For example, $\varepsilon(x^2 + 3x^7 - x^{666}) = 1+3-1 = 3$. This function ε is very important in abstract algebra, and is called the **augmentation** map.

Definition 12.2 Let V and W be vector spaces. Then a **linear transformation** (or **linear map**) $f\colon V \longrightarrow W$ is a function $f\colon V \longrightarrow W$ which satisfies the following rules:

(i) $f(v_1+v_2) = f(v_1) + f(v_2)$

(ii) $f(\lambda v) = \lambda f(v)$

for all v, v_1, v_2 in V and λ in $\mathbb{R}$.

Note that in (i) the RHS is addition in W, while the LHS is addition in V. We thus speak of f as *preserving the vector space structure*.

Examples 12.3

(a) Let $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^3$ be given by $f(x, y) = (x+2y, x, x-y)$. Then f is linear. Indeed,

(i) $$\begin{aligned} f((x, y) + (x', y')) &= f(x+x', y+y') = (x+x'+2(y+y'), x+x', x+x'-(y+y')) \\ &= (x+2y, x, x-y) + (x'+2y', x', x'-y') \\ &= f(x,y) + f(x',y'), \end{aligned}$$

and (ii) $f(\lambda(x, y)) = f(\lambda x, \lambda y) = (\lambda x+2\lambda y, \lambda x, \lambda x-\lambda y)$
$= \lambda(x+2y, x, x-y)$
$= \lambda f(x, y)$
for all (x,y) and (x',y') in $\mathbb{R}^2$ and $\lambda \in \mathbb{R}$.

(b) Let A be any $m\times n$ matrix, and let $\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ be the associated map as in Example 12.1(c). Then $\hat{A}$ is linear. Indeed, if $\boldsymbol{v}$ and $\boldsymbol{w}$ are in $\mathbb{R}^n$, then

$$\hat{A}(\boldsymbol{v}+\boldsymbol{w}) = [A.(\boldsymbol{v}+\boldsymbol{w})^t]^t = [A.(\boldsymbol{v}^t+\boldsymbol{w}^t)]^t = [A.\boldsymbol{v}^t + A.\boldsymbol{w}^t]^t$$
$$= [A.\boldsymbol{v}^t]^t + [A.\boldsymbol{w}^t]^t = \hat{A}(\boldsymbol{v}) + \hat{A}(\boldsymbol{w}),$$

and

$$\hat{A}(\lambda\boldsymbol{v}) = [A.\lambda\boldsymbol{v}^t]^t = \lambda[A\boldsymbol{v}^t]^t = \lambda\hat{A}(\boldsymbol{v}).$$

We shall see later that, essentially, all linear maps between finite dimensional spaces look like this.

(c) As a special case of the above example, let A be the matrix

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

(d) The identity map $\mathbf{1}$: $V \longrightarrow V$ described in Examples 12.1 is linear. (Homework). **Note:** If $V = \mathbb{R}^n$, then this map is of the form $\hat{I}$, where I is the $n\times n$ identity matrix.

(e) The *zero map* $\mathbf{0}$: $V \longrightarrow V$ given by $\mathbf{0}(v) = 0$, (the zero vector), for any $v \in V$, is linear. (Homework). If $V = \mathbb{R}^n$, then this map is of the form $\hat{O}$, where O is the $n\times n$ zero matrix.

(f) The **dilation** maps T: $V \longrightarrow V$ given by $T(v) = kv$ for some scalar k. If $V = \mathbb{R}^n$, then these maps are of the form $\hat{A}$, where $A = kI$.

(Illustrations of the actions of these on figures in $\mathbb{R}^n$)

(g) Let V be any n-dimensional vector space and let $\mathscr{B}$ and $\mathscr{B}'$ be any two bases for V; $\mathscr{B} = \{v_1, \ldots v_n\}$, $\mathscr{C} = \{w_1, \ldots w_n\}$. We define the **change-of-basis transformation**, $T_{\mathscr{B},\mathscr{C}}$: $V \longrightarrow V$ as follows. If $v \in V$, then one can write $v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$ for unique scalars λ_i, by above results. We now define

$$T_{\mathscr{B},\mathscr{C}}(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n) = \lambda_1 w_1 + \lambda_2 w_2 + \ldots + \lambda_n w_n.$$

That $T_{\mathscr{B},\mathscr{C}}$ is linear is in the homework.

(h) Let D: $C^1[a, b] \longrightarrow C[a, b]$ be the map given by $D(f) = f'$, (the derivative of f). Then D is a linear transformation. (By convention, $C^1[a,b]$ is the vector space of *continuously differentiable maps* on $[a, b]$. That is, the derivative exists and is continuous on the interval $[a, b]$.) Thus "differentiation is a linear operation."

(i) Define J: $C[0,1] \longrightarrow \mathbb{R}$ by $J(f) = \int_0^1 f(x)\, dx$. Then J is linear. (Exercise). For example, if f is specified by $f(x) = \sin(\pi x)$, then

$$J(f) = \int_0^1 \sin \pi x \, dx = 2/\pi.$$

Proposition 12.4 (Preservation of subtraction and 0)
If f: $V \longrightarrow W$ is linear, then:

(i) $f(v_1 - v_2) = f(v_1) - f(v_2)$ for all v_1 & $v_2 \in V$.

(ii) $f(0) = 0$ and $f(-v) = -f(v)$ for all $v \in V$.

Proof For (i),

$$f(v_1 - v_2) \underset{\substack{\uparrow \\ \text{defn of } -v}}{=} f(v_1 + (-1)v_2) \underset{\substack{\uparrow \\ \text{1st property for lin. maps}}}{=} f(v_1) + f((-1)v_2) \underset{\substack{\uparrow \\ \text{2nd property}}}{=} f(v_1) + (-1)f(v_2) \underset{\substack{\uparrow \\ \text{defn of } -f(v_2)}}{=} f(v_1) + f(v_2)$$

For (ii) First apply (i) with $v_1 = v_2 = 0$ to get first assertion, then apply first assertion and (i) with $v_1 = 0$ and $v_2 = v$ to get second assertion.

Note This should remind one of corresponding rules for derivatives and integration. In fact, these rules hold for differentiation and integration precisely because—as we have seen above—they are linear transformations.

Definitions 12.5 If f: $V \longrightarrow W$ is a linear map, then the **kernel** (*or* **null space**) **of f** is the set of vectors

$$\ker f = \{v \in V : f(v) = 0\}.$$

On the other hand, the **image of f** is the set of vectors

$$\operatorname{Im} f = \{f(v) : v \in V\}.$$

(This is just the "range" of f in the usual sense.)

Examples 12.6

(a) Let f: $V \longrightarrow W$ be the zero map. Since f annihilates everything in V, it follows that $\ker f = V$ itself. The image of f is zero.

(b) Let A be an $m \times n$ matrix, and let f be the linear map $\hat{A} : \mathbb{R}^n \longrightarrow R^m$. Then $\hat{A}(v) = 0$ iff $A.v = 0$. In other words, v is a solution to the homogeneous equation $AX = O$. Thus, $\ker \hat{A}$ is the *solution space* of $AX = O$. Turning to the image of $\hat{A}$, it consists of all w such that the matrix equation $AX = w$ has a solution (i.e.. is consistent). We'll see later that this in fact corresponds to the subspace of $\mathbb{R}^m$ generated by the columns of A.

(c) Let f: $C^1(\mathbb{R}) \longrightarrow C^0(\mathbb{R})$ be differentiation, then *ker f* is the subspace of $C^1(\mathbb{R})$ consisting of the constant maps. The Image of f is the whole of $C^0(\mathbb{R})$, since every continuous function has an antiderivative.

Theorem 12.7 (Kernel and Image are Subspaces)
If f: $V \longrightarrow W$ is linear, then $\ker f$ and Im f are subspaces of V and W respectively.

Proof Checking the two conditions for ker$f$,

x & $y \in \ker f \Rightarrow f(x) = 0$ & $f(y) = 0$
$\Rightarrow f(x+y) = f(x) + f(y) = 0 + 0 = 0$
$\Rightarrow x+y \in \ker f$.

$x \in \ker f$ and $\lambda \in \mathbb{R} \Rightarrow f(\lambda x) = \lambda f(x) = \lambda.0 = 0$
$\Rightarrow \lambda x \in \ker f$.

Checking the two conditions for Im f,

v & $w \in \operatorname{Im} f \Rightarrow v = f(x)$ & $w = f(y)$ for some x and y in V
$\Rightarrow v + w = f(x) + f(y) = f(x+y)$
$\Rightarrow v+w \in \operatorname{Im} f$.

$v \in \operatorname{Im} f$ and $\lambda \in \mathbb{R} \Rightarrow v = f(x)$, so $\lambda v = \lambda f(x) = f(\lambda x)$
$\Rightarrow \lambda v \in \operatorname{Im} f$. ➢

Exercise Set 12

Hand In (Value = 30)

1. Show that the augmentation map ε: $\mathbb{R}[x] \longrightarrow \mathbb{R}$ defined in Examples 12.1 is linear.
2. Show that the identity and zero maps on any vector space V are linear.
3. Show that the change-of-basis map $T_{\mathscr{B},\mathscr{C}}$ is linear for any vector space V ad bases $\mathscr{B}$ and $\mathscr{C}$.
4. If f: $V \longrightarrow W$ and g: $W \longrightarrow U$ are maps, one has their **composite**, $g \circ f$: $V \longrightarrow U$, given by $g \circ f(v) = g(f(v))$ for each $v \in V$ Show that, if f and g are linear, then so is $g \circ f$.
5. Say that a function f: $X \longrightarrow Y$ is **injective** ("one-to-one") $f(x) \neq f(y)$ unless $x = y$; in other words,

$$f(x) = f(y) \Rightarrow x = y.$$

(a) Show that the function f: $\mathbb{R} \longrightarrow \mathbb{R}$ given by $f(x) = 4x+5$ is injective, whereas the function g: $\mathbb{R} \longrightarrow \mathbb{R}$ given by $g(x) = x^2+1$ is not.
(b) Convince yourself that a function f: $\mathbb{R} \longrightarrow \mathbb{R}$ is injective if and only if its graph "passes the horizontal line test." (You need not hand in your conviction.)
(c) In the same vein, say that a linear transformation T: $V \longrightarrow W$ is **injective** if $T(x) = T(y)$ implies $x = y$. Prove that T is injective iff $\ker(T) = \{0\}$.

---

13. Linear Maps and Bases, Featuring: The Rank Theorem

Lemma 13.1 (Spanning the Image)
If $\{v_1, v_2, \ldots, v_n, \ldots\}$ is a basis for V and if $f\colon V \longrightarrow W$ is linear, then $\mathrm{Im}(f)$ *is* the span of $f(v_1), f(v_2), \ldots, f(v_n), \ldots$.

Proof Homework.

We now look closely at linear maps and see what makes them tick.

Proposition 13.2 (Linear Maps Determined by Action on a Basis)
If $\mathscr{B}$ is a basis for the vector space V and if $f\colon V \longrightarrow W$ is a linear map, then f is entirely determined by its values on the elements of $\mathscr{B}$. (That is, if f and g do the same thing to the basis vectors, then $f = g$.)

Proof. Let $\mathscr{B} = \{v_1, v_2, \ldots, v_n, \ldots\}$ be a basis for V. Then we can write, for any $v \in V$,

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r$$

for some scalars λ_i and some r. Thus,

$$\begin{aligned} f(v) &= f(\lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_r v_r) \\ &= f(\lambda_1 v_1) + f(\lambda_2 v_2) + \ldots + f(\lambda_r v_r) \\ &= \lambda_1 f(v_1) + \lambda_2 f(v_2) + \ldots + \lambda_r f(v_r), \end{aligned}$$

showing that f is entirely determined by the vectors

$$f(v_1),\ f(v_2),\ \ldots, f(v_r), \ldots$$

as required. ✛

Now assume that $f\colon V \longrightarrow W$ is linear with V and W *finite dimensional*.

Definition 13.3 The **rank** of f is the dimension of $\mathrm{Im}\, f$; the **nullity** of f is the dimension of $\ker f$.

Examples in class

Theorem 13.4 (Rank Theorem)
Let $f\colon V \longrightarrow W$ be linear, with V and W finite dimensional. Then

$$\mathrm{Rank}(f) + \mathrm{Nullity}(f) = \dim V.$$

Proof Let $\mathrm{Rank}(f) = r$, and $\mathrm{Nullity}(f) = k$. Then $\ker(f)$ has a basis consisting of k vectors $\{v_1, \ldots, v_k\}$, say. Since these are linearly independent, one may keep adding vectors not in its span to obtain a basis

$$\{v_1, \ldots, v_k, w_1, \ldots, w_s\}$$

of V. (Justification: Exercise Set 10 #3) I claim that the vectors $f(w_1), \ldots, f(w_s)$ form a basis for $\mathrm{Im}(f)$, and hence $r = s$. All we have to show is linear independence and the spanning property.

For linear independence, if

$$\lambda_1 f(w_1) + \ldots + \lambda_s f(w_s) = 0$$

then, since f is linear,

$$f(\lambda_1 w_1 + \ldots + \lambda_s w_s) = 0,$$

and hence

$$\lambda_1 w_1 + \ldots + \lambda_s w_s \in \ker(f).$$

Since $\ker(f)$ is spanned by $\{v_1, \ldots, v_k\}$, it follows that

$$\lambda_1 w_1 + \ldots + \lambda_s w_s = \mu_1 v_1 + \ldots + \mu_k v_k$$

for suitable scalars μ_i . But then

$$\lambda_1 w_1 + \ldots + \lambda_s w_s - \mu_1 v_1 - \ldots - \mu_k v_k = 0$$

Since $\{v_1, \ldots, v_k, w_1, \ldots, w_s\}$ is a basis for V, we must therefore have

$$\lambda_1 = \ldots = \lambda_s = \mu_1 = \ldots = \mu_k = 0.$$

But since the λ *'s* are zero, we have shown linear independence of the vectors $f(w_1), \ldots, f(w_s)$.

For the spanning property, we must show that the vectors $f(w_1), \ldots, f(w_s)$ span $\mathrm{Im}(f)$. Thus let $w \in \mathrm{Im}(f)$. Then $w = f(v)$ for some $v \in V$, by definition of the image. Since $\{v_1, \ldots, v_k, w_1, \ldots, w_s\}$ is a basis for V, we can write

$$v = \lambda_1 w_1 + \ldots + \lambda_s w_s + \mu_1 v_1 + \ldots + \mu_k v_k$$

for appropriate scalars λ_i and μ_i. Thus,

$$\begin{aligned} w = f(v) &= f(\lambda_1 w_1 + \ldots + \lambda_s w_s + \mu_1 v_1 + \ldots + \mu_k v_k) \\ &= f(\neg\lambda_1 w_1 + \ldots + \lambda_s w_s) + f(\mu_1 v_1 + \ldots + \mu_k v_k) \\ &= f(\neg\lambda_1 w_1 + \ldots + \lambda_s w_s), \end{aligned}$$

since the v's are in the kernel of f,

$$= \lambda_1 f(\neg w_1) + \ldots + \lambda_s f(w_s)$$

by linearity of f. Thus w is in the span of the $f(w_i)$'s, showing the spanning property.

We now have $s+k = \dim V$. Since $s = r$, $r+k = \dim V$, completing the proof. ✺

Proposition and Definition 13.5 (Rank of a Matrix)
Let A be an $m \times n$ matrix, and let

$$\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

be multiplication by A. Let A' be obtained from A by passage to *column reduced echelon form*. Then the non-zero rows in A' are a basis for the image of $\hat{A}$, so that

$$\mathrm{rank}(\hat{A}) = \text{number of non-zero rows after column reduction.}$$

We call this number the **column rank of the matrix** A. (The **row rank** is defined similarly.)

Proof By Lemma 13.1, the vectors $\hat{A}(e_1), \ldots, \hat{A}(e_n)$ span the image of $\hat{A}$. By direct computation,

$$\hat{A}(e_1) = 1\textit{st column of } A,$$

$$\ldots$$

$\hat{A}(e_n) = nth\ column\ of\ A.$

By Proposition 9.7, the span of the columns is not effected by column reduction, so that the nonzero columns of A' span the image of $\hat{A}$. Thus to show that they form a basis, it remains to show that they are linearly independent. But, by definition of reduced echelon form, each column has a 1 in a slot where all the others have 0's. Thus if any linear combination of them is zero, it follows that the coefficients must be zero, and we are done. ◆

The above result says how to compute the (column) rank of A. For the Nullity, we know that this is the dimension of the solution space of $AX = O$. deals with

By the Rank+Nullity theorem,

Column Rank + Dimension of Solution Space for $(AX = O) = n$.

Thus, to get the dimension of the solution space, we compute the column rank, and get

Dimension of solution space $= n -$ column rank.

Theorem 13.6 (Row Rank = Column Rank)
The row rank and column rank of any $m \times n$ matrix are the same.

Proof We show that column rank $\leq$ row rank, and then get the reverse inequality by looking at the transpose.

S'pose that the row space has dimension k, and let $\{v_1, \ldots, v_k\}$ be a basis for the row space. For each r, write the coordinates of v_r as $(v_{r_1}, v_{r_2}, \ldots, v_{r_n})$. Then each row is a linear combination of the v's. i.e..,

$$\begin{aligned} \text{row } 1 = r_1 &= c_{11}v_1 + \ldots + c_{1k}v_k \\ r_2 &= c_{21}v_1 + \ldots + c_{2k}v_k \\ &\ldots \\ r_m &= c_{m1}v_1 + \ldots + c_{mk}v_k \end{aligned}$$

Here, the rows are given by

$$\begin{aligned} r_1 &= (a_{11}, \ldots, a_{1n}), \\ &\ldots \\ r_m &= (a_{m1}, \ldots, a_{mn}), \end{aligned}$$

where a_{ij} are the entries of A. Plugging the r's into the first batch of equations and equating coordinates gives:

$$\begin{aligned} a_{1j} &= c_{11}v_{1j} + \ldots + c_{1k}v_{kj} \\ a_{2j} &= c_{21}v_{1j} + \ldots + c_{2k}v_{kj} \\ &\ldots \\ a_{mj} &= c_{m1}v_{1j} + \ldots + c_{mk}v_{kj} \end{aligned}$$

Thus

$$(a_{1j}, a_{2j}, \ldots, a_{mj}) = v_{1j}(c_{11}, \ldots, c_{m1}) + v_{1j}(c_{12}, \ldots, c_{m2}) + \ldots + v_{1j}(c_{1k}, \ldots, c_{mk}).$$

Since the LHS is the jth column of A, it follows that the jth column of A is a linear combination of the k vectors on the right. Thus the *dimension* of the column space is $\leq k$. But k is the row rank, so we get:

column rank $\leq$ row rank.

Done. ✞

It follows that to get the *rank* of A, we can either row reduce or column reduce, and will get the same answer.

Recall now that we saw, on p.55, that, if A is any $m \times n$ matrix, and if $\hat{A}$ is the linear map $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ determined by A, then

$\hat{A}\,(e_1) = $ 1st column of A,

$\hat{A}\,(e_2) = $ 2nd column of A,

. . .

$\hat{A}\,(e_n) = n$th column of A.

Thus, we can *construct* the matrix A once we know what $\hat{A}$ is required to do to the basis vectors $e_1, e_2, \ldots, e_n$ of $\mathbb{R}^n$.

Examples: Rotation, projections, shears, reflections

Exercise Set 13

Anton §5.6 #2(a), (b), (d) (e)

Hand In (Value = 15)

1. Prove Lemma 13.1

2. Use Exercise Set 11 #5 to give a new proof that row operations don't effect the (row) rank of a matrix. You are not allowed to use Proposition 9.7.

3. Show that row operations do not effect linear independence of the columns of a matrix, but may effect the span of the columns.

---

14. The Matrix of Linear Map

Here, we see that, if $f: V \longrightarrow W$ is any linear map with V and W finite dimensional, then f may be "represented" by a matrix, whether or not V and W are subspaces of Euclidean space.

Definition 14.1 Let V and W be finite dimensional vector spaces with selected bases $\mathscr{B} = \{v_1, v_2, \ldots, v_n\}$ and $\mathscr{C} = \{w_1, w_2, \ldots, w_m\}$ respectively. Also let $f: V \longrightarrow W$ be a linear transformation. Define the **matrix $[f]$ of f with respect to the bases $\mathscr{B}$ and C** as follows.

Since C is a basis for W, we have:

$$f(v_1) = a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m;$$
$$f(v_2) = a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m;$$
$$\ldots$$
$$f(v_1) = a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m,$$

for suitable unique scalars a_{ij}. We now define the matrix $[f]$ by taking its ijth entry to be a_{ij}. Therefore, by definition, the coordinates of $f(v_i)$ with respect to the basis $\{w_j\}$ is given by

$$f(v_i) = \sum_{j=1}^{n} [f]_{ji}\, w_j.$$

Notation We sometimes write $[f]_{C\mathscr{B}}$ instead of just $[f]$ if we wish to say which bases we are using.

Now let's calculate, by brute force, just what f applied to a general vector in V is. Thus, let

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \ldots + \lambda_n v_n$$

be a general vector in V. Then

$$f(v) = \lambda_1 f(v_1) + \lambda_2 f(v_2) + \ldots + \lambda_n f(v_n).$$

Substituting the above formulas for the $f(v_i)$ $'s$ gives:

$$\begin{aligned} f(v) \ = \ & \lambda_1(a_{11}w_1 + a_{21}w_2 + \ldots + a_{m1}w_m) \\ & + \lambda_2(a_{12}w_1 + a_{22}w_2 + \ldots + a_{m2}w_m) \\ & \quad \ldots \\ & + \lambda_n(a_{1n}w_1 + a_{2n}w_2 + \ldots + a_{mn}w_m). \end{aligned}$$

Rearranging terms by grouping together the terms with $w_1, w_2, \ldots$ etc. gives:

$$f(v) = \sum_{j=1}^{n} (a_{1j}\, \lambda_j)w_1 + \sum_{j=1}^{n} (a_{2j}\, \lambda_j)w_2 + \ldots + \sum_{j=1}^{n} (a_{mj}\, \lambda_j)w_m.$$

Thus the coordinates (with respect to the $w\,'s$) of $f(v)$ are given by

$$\left(\sum_{j=1}^{n} (a_{1j}\, \lambda_j)\, ,\ \sum_{j=1}^{n} (a_{2j}\, \lambda_j)\, ,\ \ldots,\ \sum_{j=1}^{n} (a_{mj}\, \lambda_j) \right)$$

From the formula for matrix multiplication, we now see that this is just the transpose of the column vector

$[f].\ (\lambda_1, \lambda_2, \ldots, \lambda_n)^t$.

Thus we can think of elements of V as column vectors $(\lambda_1, \lambda_2, \ldots, \lambda_n)$t, (i.e.. as elements of $\mathbb{R}^n$), and compute $f(v)$ by simply multiplying this column vector by $[f]$. Loosely speaking, once we are supplied with bases, V and W "are" $\mathbb{R}^n$ and $\mathbb{R}^m$ respectively, and $f\colon V \longrightarrow W$ "is" multiplication by the matrix $[f]$.

Question: Now just how do we get the matrix in an easy way?

Answer: Look at $f(v_1)$. Its coordinates, according to the original formula in Definition 14.1 are given by $(a_{11}, a_{21}, \ldots, a_{2m})$. By Definition 14.1, this is just the 1st column of $[f]$. Similarly, the *ith* column of $[f]$ is given by the coordinates of $f(v_i)$. In other words, *you get the columns of the matrix by applying f to the basis vectors*.

Examples 14.2
(a) Determine the matrix of the map $f\colon \mathbb{R}_2[x] \longrightarrow \mathbb{R}_3[x]$ given by $f(p(x)) = x\,p(x)$, with respect to the usual bases.
Answer: The usual basis for $\mathbb{R}_2[x]$ is $\{1, x, x^2\}$, and that for $\mathbb{R}_3[x]$ is $\{1, x, x^2, x^3\}$. To get the matrix $[f]$, we calculate:

$$f(1) = x = 0.1 + 1.x + 0.x^2 + 0.x^3;$$
$$f(x) = x^2 = 0.1 + 0.x + 1.x^2 + 0.x^3;$$
$$f(x^2) = x^3 = 0.1 + 0.x + 0.x^2 + 1.x^3,$$

whence

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

Note also that the columns of $[f]$ represent the image vectors x, x^2, x^3.

(b) If $\varepsilon\colon \mathbb{R}_n[x] \longrightarrow \mathbb{R}$ is the augmentation, then, with respect to the usual bases, the matrix $[\varepsilon]$ is given as follows: First, we calculate:

$$f(1) = 1$$
$$f(x) = 1$$
$$\ldots$$
$$f(x^n) = 1,$$

whence $[f]$ is the row matrix $[1\ 1\ 1\ \ldots\ 1]$ (n repetitions.)

(c) Let $\mathcal{B} = \{v_1, v_2, \ldots, v_n\}$ be any basis for $\mathbb{R}^n$, let $C = \{e_1, e_2, \ldots, e_n\}$ be the usual basis for $\mathbb{R}^n$, and let $f\colon \mathbb{R}^n \longrightarrow \mathbb{R}^n$ be the identity map. We calculate $[f]_{C\mathcal{B}}$. To do this, evaluate f on the basis elements $v_i \in \mathcal{B}$ getting:

$$f(v_1) = v_1 = (v_{11}, v_{21}, \ldots, v_{n1}) \text{ (The coordinates of v1);}$$

$f(v_2) = v_2 = (v_{12}, v_{22}, \ldots, v_{n2})$

$\ldots$

$f(v_n) = v_n = (v_{1n}, v_{2n}, \ldots, v_{nn})$.

Thus $[f]_{C\mathcal{B}}$ is the $n\times n$ matrix whose columns are the vectors $v_1, v_2, \ldots, v_n$.

(d) If $\mathcal{B}$ and C are any two bases for V, then the **transition matrix from $\mathcal{B}$ to C** is given as the matrix $[1]_{C\mathcal{B}}$, i.e.. the matrix of the identity with respect to the bases $\mathcal{B}$ and C. Thus the columns of the transition matrix are the coordinates of the elements of the first basis ($\mathcal{B}$) with respect to the second basis (C). For example, the matrix in **(c)** above is the transition matrix from $\mathcal{B}$ to the usual basis C. We also refer to $[1]_{C\mathcal{B}}$ as the **change-of-basis matrix from $\mathcal{B}$ to C**.

Theorem 14.3 (Composition of Linear Maps = Multiplication of Matrices)
If f: $V \longrightarrow W$ and g: $W \longrightarrow U$ are linear, and if $\mathcal{B}$, $\mathcal{C}$ and $\mathcal{D}$ are bases for V, W, and U respectively, then

$$[g\circ f]_{\mathcal{D}\mathcal{B}} = [g]_{\mathcal{D}\mathcal{C}}[f]_{\mathcal{C}\mathcal{B}}\,.$$

Proof One has writing $\mathcal{B} = \{b_i\}$, $\mathcal{C} = \{c_i\}$ *and* $\mathcal{D} = \{d_i\}$,

$$g\circ f(b_i) = \Sigma_j [g\circ f]_{ji}\, d_j\,,$$

by definition of the matrix for $g\circ f$. On the other hand,

$$g\circ f(b_i) = g(f(b_i)) = g(\Sigma_k [f]_{ki} c_k)$$
$$= \Sigma_k [f]_{ki} g(c_k) = \Sigma_k [f]_{ki}\, \Sigma_j [g]_{jk} d_j$$
$$= \Sigma_{k,j} [g]_{jk}\, [f]_{ki}\, d_j$$
$$= \Sigma_j ([g][f])_{ji} d_j\,.$$

Equating coefficients now gives the result. ❇

Corollary 14.4 The change-of-basis matrix $[1]_{\mathcal{B}\mathcal{C}}$ is invertible, with inverse $[1]_{\mathcal{C}\mathcal{B}}$.

Proof Since $1\circ 1 = 1$, the identity map on V, we apply the above theorem with $D = \mathcal{B}$, getting

$$[1]_{\mathcal{B}\mathcal{C}}[1]_{\mathcal{C}\mathcal{B}} = [1]_{\mathcal{B}\mathcal{B}} = I,$$

and similarly the other way 'round. ❇

Consider the following question:

Question: *Given f: $V \longrightarrow V$ with a known matrix with respect to some basis, how do we find the matrix of f with respect to some other basis?*

This question is answered by the . . .

Theorem 14.5 (Change-of-Basis)
Let f: $V \longrightarrow V$ be any linear map, let $\mathcal{B}$ be any basis of V, and let $[f]_{\mathcal{B}}$ be its matrix with respect to the basis $\mathcal{B}$. Then if $\mathcal{C}$ is any other basis, one has

$$[f]_{\mathcal{C}} = P^{-1}[f]_{\mathcal{B}}P,$$

where P is the change-of-basis matrix $[1]_{\mathcal{B}\mathcal{C}}$.

Proof Since $f = 1 \circ f \circ 1$ as a map $V \longrightarrow V$, one has, by the lemma,

$$\begin{aligned}[f]_C &= [1 \circ f \circ 1]_{\mathcal{C}} \\ &= [1]_{\mathcal{C}\mathcal{B}}[f \circ 1]_{\mathcal{B}\mathcal{C}} \\ &= [1]_{\mathcal{C}\mathcal{B}}[f]_{\mathcal{B}\mathcal{B}}[1]_{\mathcal{B}\mathcal{C}} \\ &= [1]_{\mathcal{B}\mathcal{C}}{}^{-1}[f]_{\mathcal{B}\mathcal{B}}[1]_{\mathcal{B}\mathcal{C}},\end{aligned}$$

as required. ♥

This prompts the following.

Definition 14.6 If P is any invertible $n \times n$ matrix such that $A' = P^{-1}AP$, then we say that the $n \times n$ matrices A and A' are **similar**. Thus, by Theorem 14.6, two matrices are similar if they represent the same map with respect to different bases.

Note The matrices A and A' are similar if, for *some* invertible P, $A' = P^{-1}AP$.

In order to help us compute the change-of-basis matrices $[1]_{\mathcal{B}C}$, we use the following lemma.

Lemma 14.7 If $\mathcal{B}$ and C are any bases for $\mathbb{R}^n$, then the transition matrix from $\mathcal{B}$ to C is given by

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

where P is the invertible matrix whose columns are the vectors in $\mathcal{C}$, while Q is the invertible matrix whose columns are the vectors in $\mathcal{B}$.

The proof is Exercise Set 14 #3.

Exercise Set 14
Anton §4.2 #5, 6

Hand In (Value = 15)

1. Let V and V' be vector spaces with $\mathcal{B} = \{b_1, \ldots, b_n\}$ a basis for V and $\mathcal{C} = \{c_1', \ldots, c_n'\}$ a collection of vectors in V'.

(a) Show that there exists a unique linear map T: $V \longrightarrow V'$ with the property that $T(b_i) = c_i'$ for every i.

(b) If $\mathcal{C}$ happens to be a basis, what is the matrix of this linear map?

2. Show that similarity of $n \times n$ matrices is an equivalence relation. (You might wish to refer to your erroneous proof of this fact in the first test.)

3. Show that, if $\mathscr{B}$ and $\mathscr{C}$ are any bases for $\mathbb{R}^n$, then the transition matrix from $\mathscr{B}$ to $\mathscr{C}$ is given by

$$[1]_{\mathscr{C}\mathscr{B}} = P^{-1}Q$$

where P is the invertible matrix whose columns are the vectors in $\mathscr{C}$, while Q is the invertible matrix whose columns are the vectors in $\mathscr{B}$.

4. Show that any invertible matrix P is the transition matrix associated with some change of basis $\mathscr{B} \longrightarrow \mathscr{S}$, where S is the standard basis for $\mathbb{R}^n$

15. Eigenvectors

Definitions 15.1 If f: $V \longrightarrow V$ is a linear map, then a nonzero vector $v \in V$ is called an **eigenvector of f** if $f(v) = \lambda v$ for some scalar λ. The scalar λ is called the **eigenvalue of f corresponding to the eigenvector v**. Similarly, referring to matrices, if A is an $n\times n$ matrix, then any nonzero (column) vector v is called an **eigenvector of A** if $Av = \lambda v$ for some scalar λ . λ is then called the **eigenvalue of A corresponding to the vector v.**

Examples 15.1

(a) The vector $(1, 2)^t$ is an eigenvector of $A = \begin{bmatrix} 3 & 1 \\ 2 & 4 \end{bmatrix}$ with corresponding eigenvalue 5.

(b) If A is a scalar multiple of the 2×2 identity matrix, then every vector in $\mathbb{R}^2$ is an eigenvector of A.

(c) If A is a 2×2 matrix representing rotation through a nonzero angle θ, then A has no eigenvectors.

(d) If $A = \begin{bmatrix} a & 0 \\ 0 & b \end{bmatrix}$ with $a \neq b$, then A has exactly two eigenvalues, namely a and b.

Proposition 15.2 If λ is an eigenvalue of the $n\times n$ matrix A, then λ is a solution to the equation

$$\det(A-\lambda I) = 0.$$

Noting that $\det(A-\lambda I)$ can be expressed as a polynomial expression of the unknown λ, we call this expression the **characteristic polynomial of A.**

Proof For λ to be an eigenvalue of A, it means that

$$Av = \lambda v$$

for some nonzero column vector v. That is,

$$Av = (\lambda I)v,$$

where I is the $n\times n$ identity matrix. Thus we get:

$$(A-\lambda I)v = 0.$$

This means that the matrix $(A-\lambda I)$ annihilates a nonzero vector v, so, by Theorem 5.8, it cannot be invertible. This means that $\det(A-\lambda I) = 0$, as required. ❋

Proposition 15.2 shows one how to locate all the eigenvalues: just find the roots of the characteristic polynomial.

Example 15.3

Find the eigenvalues of $A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 4 & -1 & 4 \end{bmatrix}$

This is how you get the eigen*values*. To get the eigen*vectors,* we must actually solve the homogeneous system $(A-\lambda I)X = O$ for X, once we know the eigenvalue λ to use.

Definition 15.4 The ***eigenspace of A corresponding to the eigenvalue λ*** is the subspace consisting of all the eigenvectors with eigenvalue λ. (That this is indeed a subspace is the subject of one of the homework problems.)

Example 15.5

Find a basis for the eigenspaces of $A = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$

Exercise Set 15
Anton §7.1 #1, 3, 5, 7, 8

Hand In (Value = 10)
1. Show that the set of all eigenvectors of the $n\times n$ matrix A with eigenvalue λ forms a subspace of $\mathbb{R}^n$. This subspace is called the associated **eigenspace**.
2. A non-zero $n\times n$ matrix A is **nilpotent** if $A^p = O$ for some positive integer p. Prove that the only possible eigenvalue of A.

---

16. Diagonalizing a Matrix, Where It All Comes Together

Definition 16.1 An $n\times n$ matrix A is **diagonalizable** if it is similar to an $n\times n$ diagonal matrix. That is, there exists an invertible n¿n matrix P with $P^{-1}AP$ diagonal.

Geometrically, a diagonal matrix is one which maps spheres centered at the origin into ellipsoids (also centered at the origin) with axes along the usual basis vectors. Thus diagonalizable matrices are matrices which are diagonal in the eyes of some basis $\mathscr{B}$ (possibly different from the usual one), and thus map spheres into ellipsoid-type objects whose axes are not necessarily parallel to the usual basis vectors, but are parallel instead to the vectors in the basis $\mathscr{B}$.

We now consider the question: *Under what conditions is a given $n\times n$ matrix A diagonalizable?* Here is one answer.

Theorem 16.2 (A Criterion for Diagonalizbility)
An $n\times n$ matrix is diagonalizable iff it has n linearly independent eigenvectors.

Proof $\boxed{\Rightarrow}$ First, s'pose A is diagonalizable. Then there is an invertible matrix P with $P^{-1}AP = D$, a diagonal matrix $[\lambda_1|\lambda_2| \ . \ . \ . \ \lambda_n]$. We must find the n lineary independent eigenvector. Well, take $\mathscr{B}$ to be the basis of $\mathbb{R}^n$ $\{v_1, \ . \ . \ . \ , v_n\}$ consisting of the columns in P. Then

$$AP = PD,$$

where PD is P with its *columns* multiplied by the λ_i. Equating columns now gives

$$Av_1 = \lambda_1 v_1, \ . \ . \ . \ , Av_n = \lambda_n v_n,$$

showing that all of the v_i are eigenvectors.

$\boxed{\Leftarrow}$ Conversely, if A has a basis $\mathscr{B}$ of n linearly independent eigenvectors $v_1, \ . \ . \ . \ , v_n$, then we can write

$$Av_1 = \lambda_1 v_1, \ . \ . \ . \ , Av_n = \lambda_n v_n \text{ for certain eigenvalues } \lambda_1, \ . \ . \ . \ , \lambda_n.$$

Thus, if P is the $n\times n$ matrix whose columns are the vectors $v_1, \ . \ . \ . \ , v_n$, then

$$AP = PD,$$

where D is the diagonal matrix $[\lambda_1|\lambda_2| \ . \ . \ . \ \lambda_n]$. Thus, $P^{-1}AP = D$, since P is invertible. ✯

Corollary 16.3 (How To Diagonalize A Diagonalizable Matrix)
If A is a diagonalizable matrix, and if P is the matrix whose columns are a basis of eigenvectors, then $P^{-1}AP$ is diagonal, and its diagonal entries are the eigenvalues corresponding to the eigenvectors forming the columns of P.

Example 16.4 Find an invertible matrix P which diagonalizes $A = \begin{bmatrix} 3 & -2 & 0 \\ -2 & 3 & 0 \\ 0 & 0 & 5 \end{bmatrix}$, given that we know the e/vectors $(-1, 1, 0)$, $(0, 0, 1)$ and $(1, 1, 0)$, with corresponding eigenvalues 5, 5 and 1 respectively.

(More Examples in class)

Now, we wonder, when is it the case that A does indeed have a basis of eigenvectors? Here is a partial answer.

Proposition 16.5 (If the Eigenvalues Are All Different, Then the Eigenvectors are Independent)
The set of eigenvectors corresponding to a set of *distinct* eigenvalues of A is linearly independent.

Proof S'pose $\lambda_1, \ldots, \lambda_r$ are r distinct eigenvalues whose corresponding vectors are $v_1, \ldots, v_r$. Then we must show these vectors independent. Thus s'pose not. Then we can satisfy the equation

$$\alpha_1 v_1 + \ldots + \alpha_r v_r = 0$$

with at least one $\alpha_i \neq 0$. Renumber them to make $\alpha_1 \neq 0$. Multiplying both sides by A gives:

$$\lambda_1 \alpha_1 v_1 + \ldots + \lambda_r \alpha_r v_r = 0.$$

Also, multiplying both sides by λ_r gives:

$$\lambda_r \alpha_1 v_1 + \ldots + \lambda_r \alpha_r v_r = 0$$

Subtracting the last two equations gives:

$$(\lambda_1 - \lambda_r)\alpha_1 v_1 + \ldots + (\lambda_{r-1} - \lambda_r)\alpha_{r-1} v_{r-1} + (\lambda_r - \lambda_r)\alpha_r v_r = 0.$$

Since the last term is zero, this gives:

$$(\lambda_1 - \lambda_r)\alpha_1 v_1 + \ldots + (\lambda_{r-1} - \lambda_r)\alpha_{r-1} v_{r-1} = 0.$$

Now, since $\lambda_1 \neq \lambda_r$, the coefficient of v_1 is still nonzero, and now the vectors $v_1, \ldots, v_{r-1}$ are dependent. Continue in this fashion until you end up with:

$$(\lambda_1 - \lambda_2)(\lambda_1 - \lambda_3) \ldots (\lambda_1 - \lambda_r)\, \alpha_1 v_1 = 0,$$

a contradiction. ❊

Corollary 16.6 (If The Eigenvalues Are All Different, then A is Diagonalizable) If the $n \times n$ matrix A has n distinct eigenvalues, then A is diagonalizable.

Proof Homework. ●

Some Advanced Results (which we will not prove)

Theorem 1 (Test for Diagonalizability)
The $n \times n$ matrix A is diagonalizable if and only if:
(a) its characteristic polynomial $c(x)$ factors as a product of *linear* factors, and
(b) if $(x - c_1), (x - c_2), \ldots, (x - c_r)$ are the *distinct* factors, then

$$(A - c_1 I)(A - c_2 I) \ldots (A - c_r I) = 0.$$

Theorem 2 (Diagonalizability of Symmetric Matrices)
If A is a symmetric matrix, then there is an orthogonal matrix P such that $P^{-1}AP$ is diagonal.

Exercise Set 16
Anton §7.2 #1–15 odd

Hand In (Value = 10)
1. Prove Corollary 16.6.
2. Prove that similar matrices have the same eigenvectors.

17. Classification of Finite Dimensional Vector Spaces and Other Theory

Definition 17.1 Let X and Y be sets. The function f: $X \longrightarrow Y$ is **injective** (**one-to-one**) if $f(x) = f(y) \Rightarrow x = y$.

Examples 17.2

(a) f: $\mathbb{R} \longrightarrow \mathbb{R}$; $f(x) = x^3$ is injective, whereas g: $\mathbb{R} \longrightarrow \mathbb{R}$; $g(x) = x^2$ is not,

(b) f: $\mathbb{R}[x] \longrightarrow \mathbb{R}$; $f(p(x)) = p(1)$ is linear (the augmentation map) but not injective.

(c) Let T: $\mathbb{R}^2 \longrightarrow \mathbb{R}^3$ be given by $T = \hat{A}$, with $A = \begin{bmatrix} 1 & 1 \\ 2 & 1 \\ 2 & 2 \end{bmatrix}$. Then T is injective.

The following result was proved in Exercise Set 12.

Lemma 17.3 (Criterion for Injectivity)
Let V and W be vector spaces. The linear map T: $V \longrightarrow W$ is injective iff $\ker T = \{0\}$.

Corollary 17.4
$\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is injective iff $\operatorname{rank}(A) = n$.

Proof $\hat{A}$ is injective $\Leftrightarrow \ker \hat{A} = \{0\}$

$\Leftrightarrow \dim(\ker \hat{A}) = 0$

$\Leftrightarrow \operatorname{nullity}(\hat{A}) = 0 \Leftrightarrow \operatorname{rank}(A) = n$. ◆

Definition 17.5 The function f: $X \longrightarrow Y$ is **surjective** (**onto**) if, for every $y \in Y$ there exists $x \in X$ such that $f(x) = y$. In other words: f is surjective iff $\operatorname{Im}(f) = Y$. A consequence of this is that:

Lemma 17.6
$\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^m$ is surjective iff $\operatorname{rank}(A) = m$.

Example 17.7 Is the map $\hat{A}$ in Example 17.2(c) surjective?

Definition 17.8 The function f: $X \longrightarrow Y$ is **bijective** (**one-to-one and onto**) if it is both injective and surjective.

Let A be any invertible matrix. Then, by 17.4 and 17.6, $\hat{A}$ is bijective. In fact,

Proposition 17.9
$\hat{A} : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is bijective iff A is invertible.

Definition 17.10 The functions f: $X \longrightarrow Y$ and g: $Y \longrightarrow X$ are **inverse functions** if the **composite functions** $g \circ f$: $X \longrightarrow X$ and $f \circ g$: $Y \longrightarrow Y$ are both the respective identity functions. That is,

$$g(f(x)) = x \text{ for all } x \in X,$$
and
$$f(g(y)) = y \text{ for all } y \in Y.$$

We write g as f^{-1} when this is the case.

Note If g is an inverse of f, then f is an inverse of g.

Examples 17.11
(a) exp: $\mathbb{R} \longrightarrow \mathbb{R}^+$ and $\log_e$: $\mathbb{R}^+ \longrightarrow \mathbb{R}$ are inverse functions
(b) sin: $[-\pi/2, \pi/2] \longrightarrow [-1, 1]$ and arcsin: $[-1, 1] \longrightarrow [-\pi/2. \pi/2]$ are inverse functions.
(c) f: $\mathbb{R} \longrightarrow \mathbb{R}$; $f(x) = x^2$ and g: $\mathbb{R} \longrightarrow \mathbb{R}$; $g(x) = \sqrt{x}$ are *not* inverse functions.
(d) If $B = A^{-1}$ where A and B are $n \times n$ matrices, then $\hat{A}$ and $\hat{B}$ are inverse linear functions.

Proposition 17.12
(a) The inverse of a function f: $X \longrightarrow Y$ is unique; that is, f can have at most one inverse.
(b) f: $X \longrightarrow Y$ is invertible (has an inverse) iff f is bijective.
(c) If T: $V \longrightarrow W$ is an invertible linear map, then T^{-1} is also a linear map.

Definitions 17.13 We call an invertible linear map a **linear isomorphism**. An injective linear map is a **linear monomorphism**, and a surjective linear map is a **linear epimorphism**. A linear isomorphism from a vector space V to itself is called a **linear automorphism**. Two vector spaces V and W are (**linearly**) **isomorphic** if there exists a linear isomorphism ϕ: $V \longrightarrow W$. When V and W are isomorphic, we write $V \cong W$.

Note If V and W are isomorphic, then V and W "look" the same; there is a one-to-one correspondence between their elements, and the structure is preserved by this correspondence. In other words, W is just "a copy" of V.

Examples 17.14
(a) $\mathbb{R}_n[x] \cong \mathbb{R}^{n+1}$
(b) $\mathbb{R}[x] \cong \mathbb{R}^\infty$
(c) $M[m,n] \cong \mathbb{R}^{mn}$.

Theorem 17.15 (Classification of Finite Dimensional Vector Spaces)
Two finite dimensional vector spaces V and W are isomorphic iff $\dim(V) = \dim(W)$.

It follows that every finite dimensional vector space is isomorphic with $\mathbb{R}^n$ for precisely one n. In other words, every finite dimensional vector space "looks like" $\mathbb{R}^n$ for some n.

Proposition 17.16
If f: $V \longrightarrow W$ is a linear monomorphism, then f induces a linear isomorphism $V \cong \text{Im} f$.

Extra Material: Classification of Infinite Dimensional Vector Spaces

Definition 17.17 Two sets S and T have the **same cardinality** if there exists a bijection θ: $S \longrightarrow T$. If S is any set, its **cardinality** $|A|$ is the class of all sets with the same cardinality as S. For example, the cardinality of the integers is called "aleph-0"

$$|\mathbb{Z}| = \aleph_0 ,$$

and

$$|\mathbb{R}| = \aleph_1 .$$

Theorem 17.16 (Classification of All Vector Spaces)
Two vector spaces V and W are isomorphic iff they have bases with the same cardinality.

Exercise Set 17
(Hand In Value = 25)

1. **(a)** Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be given by $T = \hat{A}$ for any $m \times n$ matrix. Show that T is injective $\Rightarrow m \geq n$.
(b) Is the condition $m \geq n$ sufficient to guarantee injectivity of T above? Prove or give a counterexample.

2. **(a)** Let T: $\mathbb{R}^n \longrightarrow \mathbb{R}^m$ be given by $T = \hat{A}$ for any $m \times n$ matrix. Show that T is surjective $\Rightarrow n \geq m$.
(b) Is the condition $n \geq m$ sufficient to guarantee surjectivity of T above? Prove or give a counterexample.

3. Let f: $V \longrightarrow W$ be any linear map, and let K' be any linear complement of $\ker f$. Show that the restriction of f to K' is a linear monomorphism.

4. Show that, if f: $V \longrightarrow W$ is any linear map, then f induces an isomorphism $K' \cong \text{Im} f$, where K' is any linear complement of $\ker f$.

☆ **A Little Extra Credit** If, without the help of others in the class, you can prove, and verbally defend your proof, of Theorem 17.16, three points will be added to your final examination score. Deadline: the week before finals.
