Math 320 Spring 2009
Part II – Linear Algebra

JWR

April 24, 2009

1 Monday February 16

1. The equation $ax = b$ has

- a unique solution $x = b/a$ if $a \neq 0$,
- no solution if $a = 0$ and $b \neq 0$,
- infinitely many solutions (namely any x) if $a = b = 0$.

2. The graph of the equation $ax + by = c$ is a line (assuming that a and b are not both zero). Two lines intersect in a unique point if they have different slopes, do not intersect at all if they have the same slope but are not the same line (i.e. if they are parallel), and intersect in infinitely many points if they are the same line. In other words, the linear system

$$a_{11}x + a_{12}y = b_1, \qquad a_{21}x + a_{22}y = b_2$$

- has a unique solution if $a_{11}a_{22} \neq a_{12}a_{21}$,
- no solution if $a_{11}a_{22} = a_{12}a_{21}$ but neither equation is a multiple of the other,
- infinitely many solutions if one equation is a multiple of the other.

3. The graph of the equation $ax + by + cz = d$ is a plane (assuming that a, b, c, are not all zero). Two planes intersect in a line unless they are parallel or identical and a line and a plane intersect in a point unless the line is parallel to the plane or lies in the plane. Hence the linear system

$$\begin{aligned} a_{11}x + a_{12}y + a_{13}z &= b_2 \\ a_{21}x + a_{22}y + a_{23}z &= b_2 \\ a_{31}x + a_{32}y + a_{33}z &= b_3 \end{aligned}$$

has either a unique solution, no solution, or infinitely many solutions.

4. A **linear system** of m equations in n unknowns $x_1, x_2, \dots, x_n$ has the form

$$\begin{array}{lll} a_{11}x_1 + a_{12}x_2 + & \cdots + a_{1n}x_n & = b_1, \\ a_{21}x_1 + a_{22}x_2 + & \cdots + a_{2n}x_n & = b_2, \\ \cdots & \cdots & \\ a_{m1}x_1 + a_{m2}x_2 + & \cdots + a_{mn}x_n & = b_m. \end{array} \qquad (\dagger)$$

The system is called **inconsistent** iff has no solution and **consistent** iff it is not inconsistent. Some authors call the system **underdetermined** iff it has infinitely many solutions i.e. if the equations do not contain enough information to determines a unique solution. The system is called **homogeneous** iff $b_1 = b_2 = \cdots b_m = 0$. A homogeneous system is always consistent because $x_1 = x_2 = \cdots = x_n = 0$ is a solution.

5. The following operations leave the set of solutions unchanged as they can be undone by another operation of the same kind.

Swap. Interchange two of the equations.

Scale. Multiply an equation by a nonzero number.

Shear. Add a multiple of one equation to a different equation.

It ie easy to see that the elementary row operations do not change the set of solutions of the system (†): each operation can be undone by another operation of the same type. Swapping two equations twice, returns to the original system, scaling a row by c and then scaling it again by c^{-1} returns to the original system, and finally adding a multiple on one row to another and then subtracting the multiple returns to the original system.

2 Wednesday February 18

6. A **matrix** is an $m \times n$ array

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

of numbers. One says that A has **size** $m \times n$ or **shape** $m \times n$ or that A has m rows and n columns. The **augmented matrix**

$$\mathbf{M} := [\mathbf{A}\ \mathbf{b}] = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} & b_m \end{bmatrix} \qquad (\ddagger)$$

represents the system of linear equations (†) in section 4. The **elementary row operations** described above to transform a system into an equivalent system may be represented as

Swap.	`M([p,q],:)=M([q,p],:)`	*Interchange the pth and qth rows.*
Scale.	`M(p,:)=c*M(p,:)`	*Multiply the pth row by c.*
Shear.	`M(p,:)=M(p,:)+c*M(q,:)`	*Add c times qth row to pth row.*

The notations used here are those of the computer language MATLAB.[1] The equal sign denotes assignment, not equality, i.e. after the command `X=Y` is executed the old value of `X` is replaced by the value of `Y`. For example, if the value of a variable `x` is 7, the effect of the command `x=x+2` is to change the value of `x` to 9.

[1] An implementation of MATLAB called OCTAVE is available free on the internet. (Do a Google search on *Octave Matlab.*) I believe it was written here at UW. A more primitive version of MATLAB called MINIMAT (short for MINIMAL MATLAB) which I wrote in 1989 is available on my website. There is a link to it on the MOODLE main page for this course. It is adequate for everything in this course. Its advantage is that it is a Java Applet and doesn't need to be downloaded to and installed on your computer. (It does require the Java Plugin for your web browser.)

Definition 7. A matrix is said to be in **echelon form** iff

(i) all zero rows (if any) occur at the bottom, and

(ii) the **leading entry** (i.e. the first nonzero entry) in any row occurs to the right of the leading entry in any row above.

It is said to be in **reduced echelon form** iff it is in echelon form and in addition

(iii) each leading entry is one, and

(iv) any other entry in the same column as a leading entry is zero.

When the matrix represents a system of linear equations, the variables corresponding to the leading entries are called **leading variables** and the other variable are called **free variables**.

Remark 8. In military lingo an echelon formation is a formation of troops, ships, aircraft, or vehicles in parallel rows with the end of each row projecting farther than the one in front. Some books use the term *row echelon form* for *echelon form* and *reduced row echelon form* or *Gauss Jordan normal form* for *reduced echelon form.*

Definition 9. Two matrices are said to be **row equivalent** iff one can be transformed to the other by elementary row operations.

Theorem 10. *If the augmented coefficient matrices (‡) of two linear systems (†) are row equivalent, then the two systems have the same solution set.*

Proof. See Paragraph 5 above. □

Remark 11. It is not hard to prove that the converse of Theorem 10 is true if the linear systems are consistent, in particular if the linear systems are homogeneous. Any two inconsistent systems have the same solution set (namely the empty set) but need not have row equivalent augmented coefficient matrices. For example, the augmented coefficient matrices

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix} \qquad \text{and} \qquad \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

corresponding to the two inconsistent systems

$$x_1 = 0, \;\; 0x_2 = 1 \qquad \text{and} \qquad 0x_1 = 1, \;\; 0x_2 = 0$$

are not row equivalent.

Theorem 12. *Every matrix is row equivalent to exactly one reduced echelon matrix.*

Proof. The **Gauss Jordan Elimination Algorithm** described in the text[2] on page 165 proves "at least one". Figure 2 shows an implementation of this algorithm in the MATLAB programming language. "At most one" is tricky. A proof appears in my book[3] on page 182 (see also page 105). □

Remark 13. Theorem 12 says that it doesn't matter which elementary row operations you apply to a matrix to transform it to reduced echelon form; you always get the same reduced echelon form.

14. Once we find an equivalent system whose augmented coefficient matrix is in reduced echelon form it is easy to say what all the solutions to the system are: the free variables can take any values and then the other variables are uniquely determined. If the last non zero row is $[0\ 0\ \cdots\ 0\ 1]$ (corresponding to an equation $0x_1 + 0x_2 + \cdots + 0x_n = 1$) then the system is inconsistent. For example, the system corresponding to the reduced echelon form

$$\begin{bmatrix} 0 & 1 & 5 & 0 & 6 & 7 \\ 0 & 0 & 0 & 1 & 8 & 5 \end{bmatrix}$$

is

$$\begin{array}{lllll} x_2 & +5x_3 & & +6x_5 & = 7 \\ & & x_4 & +8x_5 & = 5. \end{array}$$

The free variables are x_1, x_3, x_5 and the general solution is

$$x_2 = 7 - 5x_3 - 6x_5, \qquad x_4 = 5 - 8x_5$$

where x_1, x_3, x_5 are arbitrary.

3 Friday February 20

Now we define the operations of **matrix algebra**. This algebra is very useful for (among other things) manipulating linear systems. The crucial point is that all the usual laws of arithmetic hold except for the commutative law.

[2] Edwards & Penny: *Differential Equations & Linear Algebra*, 2nd ed.
[3] Robbin: *Matrix Algebra Using* MINIMAL MATLAB

Figure 1: Reduced Echelon (Gauss Jordan) Form

```
function [R, lead, free] = gj(A)
 [m n] = size(A);
 R=A; lead=zeros(1,0); free=zeros(1,0);
 r = 0;  % rank of first k columns
 for k=1:n
    if r==m, free=[free, k:n]; return; end
    [y,h] = max(abs(R(r+1:m, k))); h=r+h; % (*)
    if  (y < 1.0E-9) % (i.e if y == 0)
         free = [free, k];
    else
        lead = [lead, k]; r=r+1;
        R([r h],:) = R([h r],:);    % swap
        R(r,:) = R(r,:)/R(r,k);     % scale
        for i = [1:r-1,r+1:m]       % shear
           R(i,:) = R(i,:) - R(i,k)*R(r,:);
        end
   end % if
end  % for
```

(The effect of the line marked (*) in the program is to test that the column being considered contains a leading entry. The swap means that the subsequent rescaling is by the largest possible entry; this minimizes the relative roundoff error in the calculation.)

Definition 15. Two matrices are **equal** iff they have the same size (SIZE MATTERS!) and corresponding entries are equal, i.e. $\mathbf{A} = \mathbf{B}$ iff $\mathbf{A}$ and $\mathbf{B}$ are both $m \times n$ and

$$\text{entry}_{ij}(\mathbf{A}) = \text{entry}_{ij}(\mathbf{B})$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The text sometimes writes $\mathbf{A} = [a_{ij}]$ to indicate that $a_{ij} = \text{entry}_{ij}(\mathbf{A})$.

Definition 16. Matrix Addition. Two matrices may be added only if they are the same size; addition is performed elementwise, i.e if $\mathbf{A}$ and $\mathbf{B}$ are $m \times n$ matrices then

$$\text{entry}_{ij}(\mathbf{A} + \mathbf{B}) := \text{entry}_{ij}(\mathbf{A}) + \text{entry}_{ij}(\mathbf{B})$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$. The **zero matrix** (of whatever size) is the matrix whose entries are all zero and is denoted by $\mathbf{0}$. Subtraction is defined by

$$\mathbf{A} - \mathbf{B} := \mathbf{A} + (-\mathbf{B}), \qquad -\mathbf{B} := (-1)\mathbf{B}.$$

Definition 17. Scalar Multiplication. A matrix can be multiplied by a number (scalar); every entry is multiplied by that number, i.e. if $\mathbf{A}$ is an $m \times n$ matrix and c is a number, then

$$\text{entry}_{ij}(c\mathbf{A}) := c\,\text{entry}_{ij}(\mathbf{A})$$

for $i = 1, 2, \ldots, m$ and $j = 1, 2, \ldots, n$.

18. The operations of matrix addition and scalar multiplication satisfy the following laws:

$(\mathbf{A} + \mathbf{B}) + \mathbf{C} = \mathbf{A} + (\mathbf{B} + \mathbf{C})$.	(Additive Associative Law)
$\mathbf{A} + \mathbf{B} = \mathbf{B} + \mathbf{A}$.	(Additive Commutative Law)
$\mathbf{A} + \mathbf{0} = \mathbf{A}$.	(Additive Identity)
$\mathbf{A} + (-\mathbf{A}) = \mathbf{0}$.	(Additive Inverse)
$c(\mathbf{A} + \mathbf{B}) = c\mathbf{A} + c\mathbf{B}$,	(Distributive Laws)
$(b + c)\mathbf{A} = b\mathbf{A} + c\mathbf{A}$.	
$(bc)\mathbf{A} = b\big(c\mathbf{A}\big)$.	(Scalar Asscociative Law)
$1\mathbf{A} = \mathbf{A}$.	(Scalar Unit)
$0\mathbf{A} = \mathbf{0}$, $c\mathbf{0} = \mathbf{0}$	(Multiplication by Zero).

In terms of lingo we will meet later in the semester these laws say that the set of all $m \times n$ matrices form a **vector space**.

Definition 19. The product of the matrix $\mathbf{A}$ and the matrix $\mathbf{B}$ is defined only if the number of columns in $\mathbf{A}$ is the same as the number or rows in $\mathbf{B}$, and it that case the product $\mathbf{AB}$ is defined by

$$\text{entry}_{ik}(\mathbf{AB}) = \sum_{j=1}^{n} \text{entry}_{ij}(\mathbf{A})\text{entry}_{jk}(\mathbf{B})$$

for $i = 1, 2, \ldots, m$ and $k = 1, 2, \ldots, p$ where $\mathbf{A}$ is $m \times n$ and $\mathbf{B}$ is $n \times p$. Note that the ith row of $\mathbf{A}$ is a $1 \times n$ matrix, the kth column of $\mathbf{B}$ is a $n \times 1$ matrix, and

$$\text{entry}_{ik}(\mathbf{AB}) = \text{row}_i(\mathbf{A})\text{column}_k(\mathbf{B}).$$

20. With the notations

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}, \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix},$$

the linear system (†) of section 4 may be succinctly written

$$\mathbf{Ax} = \mathbf{b}.$$

21. A **square matrix** is one with the same number of rows as columns, i.e of size $n \times n$. A **diagonal matrix** is a square matrix $\mathbf{D}$ of form

$$\mathbf{D} = \begin{bmatrix} d_1 & 0 & \cdots & 0 \\ 0 & d_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & d_n \end{bmatrix}$$

i.e. all the nonzero entries are on the **diagonal**. The **identity matrix** is the square matrix whose diagonal entries are all 1. We denote the identity matrix (of any size) by $\mathbf{I}$. The jth column of the identity matrix is denoted by $\mathbf{e}_j$ and is called the jth **basic unit vector**. Thus

$$\mathbf{I} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}.$$

22. The matrix operations satisfy the following laws:

$(\mathbf{AB})\mathbf{C}) = \mathbf{A}(\mathbf{BC})$, $(a\mathbf{B})\mathbf{C}) = a(\mathbf{BC})$.	(Asscociative Laws)
$\mathbf{C}(\mathbf{A}+\mathbf{B}) = \mathbf{CA}+\mathbf{CB}$.	(Left Distributive Law)
$(\mathbf{A}+\mathbf{B})\mathbf{C} = \mathbf{AC}+\mathbf{BC}$.	(Right Distributive Law)
$\mathbf{IA} = \mathbf{A}$, $\mathbf{AI} = \mathbf{A}$.	(Multiplicative Identity)
$\mathbf{0A} = \mathbf{0}$, $\mathbf{A0} = \mathbf{0}$.	(Multiplication by Zero)

23. The commutative law for matrix multiplication is in general false. Two matrices $\mathbf{A}$ and $\mathbf{B}$ are said to **commute** if $\mathbf{AB} = \mathbf{BA}$. This can only happen when both $\mathbf{A}$ and $\mathbf{B}$ are square and when they have the same size, but even then it can be false. For example,

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{but} \qquad \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

4 Monday February 23

Definition 24. An **elementary matrix** is a matrix which results from the identity matrix by performing a single elementary row operation.

Theorem 25 (Elementary Matrices and Row Operations)**.** *Let* $\mathbf{A}$ *be an an* $m \times n$ *matrix and* $\mathbf{E}$ *be an* $m \times m$ *elementary matrix. Then the product* $\mathbf{EA}$ *is equal to the matrix which results from applying to* $\mathbf{A}$ *the same elementary row operation as was used to produce* $\mathbf{E}$ *from* $\mathbf{I}$.

26. Suppose a matrix $\mathbf{A}$ is transformed to a matrix $\mathbf{R}$ by elementary row operations, i.e.

$$\mathbf{R} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1\mathbf{A}$$

where each $\mathbf{E}_j$ is elementary. Thus $\mathbf{R} = \mathbf{MA}$ where $\mathbf{M} = \mathbf{E}_k \cdots \mathbf{E}_2\mathbf{E}_1$. Because of the general rule

$$\mathbf{E}\begin{bmatrix} \mathbf{A} & \mathbf{B} \end{bmatrix} = \begin{bmatrix} \mathbf{EA} & \mathbf{EB} \end{bmatrix}$$

(we might say that matrix multiplication distributes over concatenation) we can find $\mathbf{M}$ via the formula

$$\mathbf{M}\begin{bmatrix} \mathbf{A} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{MA} & \mathbf{MI} \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{M} \end{bmatrix}.$$

The MATLAB program shown in Figure 4 implements this algorithm.

Figure 2: Reduced Echelon Form and Multiplier

```
function [M, R] = gjm(A)
     [m,n] = size(A);
     RaM = gj([A eye(m)]);
     R = RaM(:,1:n);
     M = RaM(:,n+1:n+m);
```

Definition 27. A matrix $\mathbf{B}$ is called a **right inverse** to the matrix $\mathbf{A}$ iff $\mathbf{AB} = \mathbf{I}$. A matrix $\mathbf{C}$ is called a **left inverse** to the matrix $\mathbf{A}$ iff $\mathbf{CA} = \mathbf{I}$. The matrix $\mathbf{A}$ is called **invertible** iff it has a left inverse and a right inverse.

Theorem 28 (Uniqueness of the Inverse)**.** *If a matrix has both a left inverse and a right inverse then they are equal. Hence if* $\mathbf{A}$ *is invertible there is exactly one matrix denoted* $\mathbf{A}^{-1}$ *such that*

$$\mathbf{A}\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{A} = \mathbf{I}.$$

Proof. $\mathbf{C} = \mathbf{CI} = \mathbf{A}(\mathbf{AB}) = (\mathbf{CA})\mathbf{B} = \mathbf{IB} = \mathbf{B}.$ □

Definition 29. The matrix $\mathbf{A}^{-1}$ is called *the (not an)* **inverse** of $\mathbf{A}$.

Remark 30. The example

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

shows that a nonsquare matrix can have a one-sided inverse. Since

$$\begin{bmatrix} 1 & 0 & c_{13} \\ 0 & 1 & c_{23} \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

we see that left inverses are *not* unique. Since

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ b_{31} & b_{32} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

we see that right inverses are *not* unique. Theorem 28 says two sided inverses *are* unique. Below we will prove that an invertible matrix must be square.

5 Wednesday February 25

31. Here is what usually (but not always) happens when we transform an $m\times n$ matrix $\mathbf{A}$ to a matrix $\mathbf{R}$ in reduced echelon form. (The phrase "usually but not always" means that this is what happens if the matrix is chosen at random using (say) the MATLAB command `A=rand(m,n)`.)

Case 1: (More columns than rows). If $m < n$ then (usually but not always) matrix $\mathbf{R}$ has no zero rows on the bottom and has the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix}$$

where $\mathbf{I}$ denotes the $m \times m$ identity matrix. In this case the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has nontrivial (i.e. nonzero) solutions, the inhomogeneous system $\mathbf{Ax} = \mathbf{b}$ is consistent for every $\mathbf{b}$, and both the matrix $\mathbf{A}$ and the matrix $\mathbf{R}$ have infinitely many right inverses. In this case the last $n - m$ variables are the free variables. The homogeneous system $\mathbf{Ax} = \mathbf{0}$ takes the form

$$\mathbf{Rx} = \begin{bmatrix} \mathbf{I} & \mathbf{F} \end{bmatrix} \begin{bmatrix} \mathbf{x}' \\ \mathbf{x}'' \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \qquad \mathbf{x}' = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{bmatrix}, \quad \mathbf{x}'' = \begin{bmatrix} x_{m+1} \\ x_{m+2} \\ \vdots \\ x_n \end{bmatrix}$$

and the general solution of the homogeneous system is given by

$$\mathbf{x}' = -\mathbf{F}\mathbf{x}''.$$

(The free variables $\mathbf{x}''$ determines the other variables $\mathbf{x}'$.)

Case 2: (More rows than columns). If $n < m$ then (usually but not always) the matrix $\mathbf{R}$ has $m - n$ zero rows on the bottom and has the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix}$$

where $\mathbf{I}$ denotes the $n \times n$ identity matrix. In this case the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has no nontrivial (i.e. nonzero) solutions, the inhomogeneous system $\mathbf{Ax} = \mathbf{b}$ is inconsistent for infinitely $\mathbf{b}$, and both the matrix $\mathbf{A}$ and the matrix $\mathbf{R}$ have infinitely many left inverses. The system $\mathbf{Ax} = \mathbf{b}$ may be written as $\mathbf{Rx} = \mathbf{MAx} = \mathbf{Mb}$ or

$$\begin{bmatrix} \mathbf{I} \\ \mathbf{0} \end{bmatrix} \mathbf{x} = \begin{bmatrix} \mathbf{x} \\ \mathbf{0} \end{bmatrix} = \mathbf{Mb} = \begin{bmatrix} \mathbf{b}' \\ \mathbf{b}'' \end{bmatrix}, \quad \mathbf{b}' = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}, \quad \mathbf{b}'' = \begin{bmatrix} b_{n+1} \\ b_{n+2} \\ \vdots \\ b_m \end{bmatrix}$$

which is inconsistent unless $\mathbf{b}'' = \mathbf{0}$.

Case 3: (Square matrix). If $n = m$ then (usually but not always) the matrix $\mathbf{R}$ is the identity and the matrix $\mathbf{A}$ is invertible. In this case the inverse matrix $\mathbf{A}^{-1}$ is the multiplier $\mathbf{M}$ found by the algorithm in paragraph 26 and figure 4. In this case the homogeneous system $\mathbf{Ax} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$ and the inhomogeneous system $\mathbf{Ax} = \mathbf{b}$ has the unique solution $\mathbf{x} = \mathbf{A}^{-1}\mathbf{b}$.

Theorem 32 (Invertible Matrices and Elementary Matrices)**.** *Elementary matrices are invertible. A matrix is invertible if and only if it is row equivalent to the identity, i.e. if and only if if it is a product of elementary matrices.*

Proof. See Case(3) of paragraph 31 and paragraph 35 below. □

Theorem 33 (Algebra of Inverse Matrices)**.** *The invertible matrices satisfy the following three properties:*

1. *The identity matrix* $\mathbf{I}$ *is invertible and*

$$\mathbf{I}^{-1} = \mathbf{I}.$$

2. *The inverse of an invertible matrix* $\mathbf{A}$ *is invertible and*

$$\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}.$$

3. *The product* $\mathbf{AB}$ *of two invertible matrices is invertible, and*

$$\left(\mathbf{AB}\right)^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}.$$

Proof. That $\mathbf{I}^{-1} = \mathbf{I}$ follows from $\mathbf{IC} = \mathbf{CI} = \mathbf{I}$ if $\mathbf{C} = \mathbf{I}$. That $\left(\mathbf{A}^{-1}\right)^{-1} = \mathbf{A}$ follows from $\mathbf{CA} = \mathbf{AC} = \mathbf{I}$ if $\mathbf{C} = \mathbf{A}^{-1}$. To prove $\left(\mathbf{AB}\right)^{-1} = \mathbf{B}^{-1}\mathbf{A}^{-1}$ let $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$. Then $\mathbf{C}(\mathbf{AB}) = \mathbf{B}^{-1}\mathbf{A}^{-1}\mathbf{AB} = \mathbf{B}^{-1}\mathbf{IB} = \mathbf{B}^{-1}\mathbf{B} = \mathbf{I}$ and $(\mathbf{AB})\mathbf{C} = \mathbf{ABB}^{-1}\mathbf{A}^{-1} = \mathbf{AIA}^{-1} = \mathbf{AA}^{-1} = \mathbf{I}$. □

Remark 34. Note the structure of the last proof. We are proving "*If P, then Q*" where "*if P*" is the statement "*if* $\mathbf{A}$ *and* $\mathbf{B}$ *are invertible*" and "*then Q*" is the statement "*then* $\mathbf{AB}$ *is invertible*". The first step is "*Assume that* $\mathbf{A}$ *and* $\mathbf{B}$ *are invertible.*" The second step is "*Let* $\mathbf{C} = \mathbf{B}^{-1}\mathbf{A}^{-1}$." Then there is some calculation. The penultimate step is "*Therefore* $\mathbf{C}(\mathbf{AB}) = (\mathbf{AB})\mathbf{C} = \mathbf{I}$" and the last step is "*Therefore* $\mathbf{AB}$ *is invertible*". Each step is either a hypothesis (like the first step) or introduces notation (like the second step) or follows from earlier steps. The last step follows from the penultimate step by the definition of what it means for a matrix to be invertible.

35. The algorithm in paragraph 26 can be used to compute the inverse $\mathbf{A}^{-1}$ of an invertible matrix $\mathbf{A}$ as follows. We form the $n \times 2n$ matrix $\begin{bmatrix} \mathbf{A} & \mathbf{I} \end{bmatrix}$. Performing elementary row operations produces a sequence

$$\begin{bmatrix} \mathbf{A} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}_0 & \mathbf{B}_0 \end{bmatrix}, \begin{bmatrix} \mathbf{A}_1 & \mathbf{B}_1 \end{bmatrix}, \cdots \begin{bmatrix} \mathbf{A}_m & \mathbf{B}_m \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{M} \end{bmatrix},$$

where each matrix in the sequence is obtained from the previous one by multiplication by an elementary matrix

$$\begin{bmatrix} \mathbf{A}_{k+1} & \mathbf{B}_{k+1} \end{bmatrix} = \mathbf{E}_k \begin{bmatrix} \mathbf{A}_k & \mathbf{B}_k \end{bmatrix} = \begin{bmatrix} \mathbf{E}_k\mathbf{A}_k & \mathbf{E}\mathbf{B}_k \end{bmatrix}.$$

Hence there is an **invariant relation**

$$\mathbf{A}_{k+1}^{-1}\mathbf{B}_{k+1} = (E_k\mathbf{A}_k)^{-1}(\mathbf{E}\mathbf{B}_k) = \mathbf{A}_k^{-1}\mathbf{E}^{-1}\mathbf{E}\mathbf{B}_k = \mathbf{A}_k^{-1}\mathbf{B}_k,$$

i.e. the matrix $\mathbf{A}_k^{-1}\mathbf{B}_k$ doesn't change during the algorithm. Hence

$$\mathbf{A}^{-1} = \mathbf{A}^{-1}\mathbf{I} = \mathbf{A}_0^{-1}\mathbf{B}_0 = \mathbf{A}_m^{-1}\mathbf{B}_m = \mathbf{I}^{-1}\mathbf{M} = \mathbf{M}.$$

This proves that the algorithm computes the inverse $\mathbf{A}^{-1}$ when $\mathbf{A}$ is invertible. See Case 3 of paragraph 31.

36. If n is an integer and $\mathbf{A}$ is a square matrix, we define

$$\mathbf{A}^n := \underbrace{\mathbf{A}\mathbf{A}\cdots\mathbf{A}}_{n}$$

for $n \geq 0$ with

$$\mathbf{A}^{-n} := \left(\mathbf{A}^{-1}\right)^n.$$

The power laws

$$\mathbf{A}^{m+n} = \mathbf{A}^m\mathbf{A}^n, \qquad \mathbf{A}^0 = \mathbf{I}$$

follow from these definitions and the associative law.

6 Friday February 27 and Monday March 2

37. In this section $\mathbf{A}$ denotes an $n \times n$ matrix and $\mathbf{a}_j$ denotes the jth column of $\mathbf{A}$. We indicate this by writing

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}.$$

We also use the notation $\mathbf{e}_j$ for the jth column of the identity matrix $\mathbf{I}$:

$$\mathbf{I} = \begin{bmatrix} \mathbf{e}_1 & \mathbf{e}_2 & \cdots & \mathbf{e}_n \end{bmatrix}.$$

Theorem 38. *There is a unique function called the* **determinant**[4] *which assigns a number* $\det(\mathbf{A})$ *to each square matrix* $\mathbf{A}$ *and has the following properties.*

(1) *The determinant of the identity matrix is one:*

$$\det(\mathbf{I}) = 1.$$

(2) *The determinant is additive in each column:*

$$\det(\begin{bmatrix} \cdots & \mathbf{a}_j' + \mathbf{a}_j'' & \cdots \end{bmatrix}) = \det(\begin{bmatrix} \cdots & \mathbf{a}_j' & \cdots \end{bmatrix}) + \det(\begin{bmatrix} \cdots & \mathbf{a}_j'' & \cdots \end{bmatrix})$$

(3) *Rescaling a column multiplies the determinant by the same factor:*

$$\det(\begin{bmatrix} \mathbf{a}_1 & \cdots & c\mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}) = c\det(\begin{bmatrix} \mathbf{a}_1 & \cdots & \mathbf{a}_j & \cdots & \mathbf{a}_n \end{bmatrix}).$$

(4) *The determinant is skew symmetric in the columns: Interchanging two columns reverses the sign:*

$$\det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots \mathbf{a}_j & \cdots \end{bmatrix}) = -\det(\begin{bmatrix} \cdots & \mathbf{a}_j & \cdots \mathbf{a}_i & \cdots \end{bmatrix}).$$

Lemma 39. *The following properties of the determinant function follow from properties (2-4)*

(5) *Adding a multiple of one column to a different column leaves the determinant unchanged:*

$$\det(\begin{bmatrix} \cdots & \mathbf{a}_i + c\mathbf{a}_j & \cdots \end{bmatrix}) = \det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots \end{bmatrix}) \qquad (i \neq j)$$

[4] The text uses the notation $|\mathbf{A}|$ where I have written $\det(\mathbf{A})$.

(6) *If a matrix has two identical columns its determinant is zero:*

$$i \neq j,\ \mathbf{a}_i = \mathbf{a}_j \implies \det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots & \mathbf{a}_j & \cdots \end{bmatrix}) = 0.$$

Proof. Item (6) is easy: interchanging the two columns leaves the matrix unchanged (because the columns are identical) and reverses the sign by item (4). To prove (5)

$$\begin{aligned} \det(\begin{bmatrix} \cdots & \mathbf{a}_i + c\mathbf{a}_j & \cdots \end{bmatrix}) &= \det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots \end{bmatrix}) + \det(\begin{bmatrix} \cdots & c\mathbf{a}_j & \cdots \end{bmatrix}) \\ &= \det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots \end{bmatrix}) + c\det(\begin{bmatrix} \cdots & \mathbf{a}_j & \cdots \end{bmatrix}) \\ &= \det(\begin{bmatrix} \cdots & \mathbf{a}_i & \cdots \end{bmatrix}) \end{aligned}$$

by (2), (3), and (6) respectively. □

Remark 40. The theorem defines the determinant implicitly by saying that there is only one function satisfying the properties (1-4). The text gives an inductive definition of the determinant on page 201. "Inductive" means that the determinant of an $n \times n$ matrix is defined in terms of other determinants (called **minors**) of certain $(n-1) \times (n-1)$ matrices. Other definitions are given in other textbooks. We won't prove Theorem 38 but will instead show how it gives an algorithm for computing the determinant. (This essentially proves the uniqueness part of Theorem 38.)

Example 41. The determinant of a 2×2 matrix is given by

$$\det\left(\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}\right) = a_{11}a_{22} - a_{12}a_{21}.$$

The determinant of a 3×3 matrix is given by

$$\begin{aligned} \det\left(\begin{bmatrix} a_{11} & a_{12} & a_{31} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix}\right) = \quad & a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} \\ & - a_{12}a_{21}a_{33} - a_{11}a_{23}a_{32} - a_{13}a_{22}a_{31}. \end{aligned}$$

The student should check that (with these definitions) the properties of the determinant listed in Theorem 38 hold.

Theorem 42 (Elementary Matrices and Column Operations)**.** *Let* $\mathbf{A}$ *be an an* $m \times n$ *matrix and* $\mathbf{E}$ *be an* $n \times n$ *elementary matrix. Then the product*

AE *is equal to the matrix which results from applying to* **A** *the same elementary column operation as was used to produce* **E** *from* **I**. *(The* **elementary column operations** *are swapping two columns, rescaling a column by a nonzero factor, and adding a multiple of one column to another.)*

Proof. This is just like Theorem 25. (The student should write out the proof for 2×2 matrices.) □

Theorem 43. *If* **E** *is an elementary matrix, and* **A** *is a square matrix of the same size, then the determinant of the product* **AE** *is given by*

(Swap) $\det(\mathbf{AE}) = -\det(\mathbf{A})$ *if (right multiplication by)* **E** *swaps two columns;*

(Scale) $\det(\mathbf{AE}) = c\det(\mathbf{A})$ *if* **E** *rescales a column by c;*

(Shear) $\det(\mathbf{AE}) = \det(\mathbf{A})$ *if* **E** *adds a multiple of one column to another.*

Proof. These are properties (2-4) in Theorem 38. □

Theorem 44. *The determinant of a product is the product of the determinants:*

$$\det(\mathbf{AB}) = \det(\mathbf{A})\det(\mathbf{B}).$$

Hence a matrix is invertible if and only if its determinant is nonzero and $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$.

Proof. An invertible matrix is a product of elementary matrices, so this follows from Theorem 43 if **A** and **B** are invertible. Just as a noninvertible square matrix can be transformed to a matrix with a row of zeros by elementary row operations so also a noninvertible square matrix can be transformed to a matrix with a column of zeros by elementary column operations. A matrix with a column of zeros has determinant zero because of part (3) of Theorem 38: multiplying the zero column by 2 leaves the matrix unchanged and mutiplies the determinanat by 2 so the determinant must be zero. Hence the determinant is zero if either **A** or **B** (and hence also **AB**) in not invertible. The formula $\det(\mathbf{A}^{-1}) = \det(\mathbf{A})^{-1}$ follows as $\det(\mathbf{A}^{-1})\det(\mathbf{A}) = \det(\mathbf{A}^{-1}\mathbf{A}) = \det(\mathbf{I}) = 1$. The fact that a matrix is invertible if and only if its determinant is nonzero follows form the facts that an invertible matrix is a product of elementary matrices (Theorem 32), the determinant of an elementary matrix is not zero (by Theorem 43 with $\mathbf{A} = \mathbf{I}$), and the determinant of a matrix with a zero column is zero. □

Remark 45. The text contains a formula for the inverse of a matrix in terms of determinants (the **transposed matrix of cofactors**) and a related formula (**Cramer's Rule**) for the solution of the inhomogeneous system $\mathbf{Ax} = \mathbf{b}$ where $\mathbf{A}$ is invertible. We will skip this, except that the student should memorize the formula

$$\mathbf{A}^{-1} = \frac{1}{\det(\mathbf{A})} \begin{bmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{bmatrix}$$

for the inverse of the 2×2 matrix

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}.$$

Corollary 46. *If* $\mathbf{E}$ *is an elementary matrix, and* $\mathbf{A}$ *is a square matrix of the same size, then the determinant of the product* $\mathbf{EA}$ *is given by*

(Swap) $\det(\mathbf{EA}) = -\det(\mathbf{A})$ *if (left multiplication by)* $\mathbf{E}$ *swaps two rows;*

(Scale) $\det(\mathbf{EA}) = c\det(\mathbf{A})$ *if* $\mathbf{E}$ *rescales a row by* c*;*

(Shear) $\det(\mathbf{EA}) = \det(\mathbf{A})$ *if* $\mathbf{E}$ *adds a multiple of one row to another.*

47. Figure 6 shows a MATLAB program which uses this corollary to compute the determinant at the same time as it computes the reduced echelon form. The algorithm can be understood as follows. If $\mathbf{E}$ is an elementary matrix then

$$\det(\mathbf{EA}) = c\det(\mathbf{A})$$

where c is the scale factor if multiplication by $\mathbf{E}$ rescales a row, $c = -1$ if multiplication by $\mathbf{E}$ swaps two rows, and $c = 1$ if multiplication by $\mathbf{E}$ subtracts a multiple of one row from another. We initialize a variable d to 1 and as we transform $\mathbf{A}$ we update d so that the relation $d\det(\mathbf{A}) = k$ always holds with k constant, i.e. $k = \det(\mathbf{A})$. (This is called an **invariant relation** in computer science lingo.) Thus when we rescale a row by c^{-1} we replace d by dc, when we swap two rows we replace d by $-d$, and when when we subtract one row from another we leave d unchanged. (The matrix $\mathbf{A}$ changes but k does not.) At the end we have replaced $\mathbf{A}$ by $\mathbf{I}$ so $d\det(\mathbf{I}) = k$ so $d = k = \det(\mathbf{A})$.

Figure 3: Computing the Determinant and Row Operations

```
function d = det(A)
    % invariant relation d*det(A) = constant
    [m n] = size(A);  d=1;
    for k=1:n
       [y,h] = max(abs(A(k:m, k))); h=k-1+h;
       if  y < 1.0E-9  % (i.e if y == 0)
           d=0; return
       else
           if (k~=h)
              A([k h],:) = A([h k],:); % swap
              d=-d;
           end
           c = A(k,k);
           A(k,:) = A(k,:)/c;         % scale
           d=c*d;
           for i =  k+1:m             % shear
              A(i,:) = A(i,:) - A(i,k)*A(k,:);
           end
      end % if
   end  % for
```

7 Monday March 2

Definition 48. The **transpose** of an $m \times n$ matrix $\mathbf{A}$ is the $n \times m$ matrix $\mathbf{A}^\top$ defined by

$$\text{entry}_{ij}(\mathbf{A}^\top) = \text{entry}_{ji}(A)$$

for $i = 1, \ldots, n$, $j = 1, \ldots, m$.

49. The following properties of the transpose operation (see page 206 of the text) are easy to prove:

(i) $(\mathbf{A}^\top)^\top = \mathbf{A}$;

(ii) $(\mathbf{A} + \mathbf{B})^\top = \mathbf{A}^\top + \mathbf{B}^\top$;

(iii) $(c\mathbf{A}^\top) = c\mathbf{A}^\top$;

(iv) $(\mathbf{AB})^\top = \mathbf{B}^\top \mathbf{A}^\top$.

For example, to prove (iv)

$$\begin{aligned}
&\text{entry}_{ij}((\mathbf{AB})^\top) = \text{entry}_{ji}(\mathbf{AB}) = \sum_k \text{entry}_{jk}(\mathbf{A})\text{entry}_{ki}(\mathbf{B}) \\
&= \sum_k \text{entry}_{ki}(\mathbf{B})\text{entry}_{jk}(\mathbf{A}) = \sum_k \text{entry}_{ik}(\mathbf{B}^\top)\text{entry}_{kj}(\mathbf{A}^\top) \\
&= \text{entry}_{ij}(\mathbf{B}^\top \mathbf{A}^\top).
\end{aligned}$$

Note also that the transpose of an elementary matrix is again an elementary matrix. For example.

$$\begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}^\top = \begin{bmatrix} 1 & 0 \\ c & 1 \end{bmatrix}, \qquad \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}^\top = \begin{bmatrix} c & 0 \\ 0 & 1 \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}^\top = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Finally a matrix $\mathbf{A}$ is invertible if and only if its transpose $\mathbf{A}^\top$ is invertible (because $\mathbf{BA} = \mathbf{AB} = \mathbf{I} \implies \mathbf{A}^\top\mathbf{B}^\top = \mathbf{B}^\top\mathbf{A}^\top = \mathbf{I}^\top = \mathbf{I}$) and the inverse of the transpose is the transpose of the inverse:

$$\left(\mathbf{A}^\top\right)^{-1} = \left(\mathbf{A}^{-1}\right)^\top.$$

Remark 50. The text (see page 235) does not distinguish $\mathbb{R}^n$ and $\mathbb{R}^{n\times 1}$ and sometimes uses parentheses in place of square brackets for typographical reasons. It also uses the transpose notation for the same purpose so

$$\mathbf{x} = (x_1, x_2, \dots, x_n) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Theorem 51. *A matrix and its transpose have the same determinant:*

$$\det(\mathbf{A}^{\mathsf{T}}) = \det(\mathbf{A}).$$

Proof. The theorem is true for elementary matrices and every invertible matrix is a product of elementary matrices. Hence it holds for invertible matrices by Theorem 44. If $\mathbf{A}$ is not invertible then $\det(\mathbf{A}^{\mathsf{T}}) = 0$ and $\det(\mathbf{A}) = 0$ so again $\det(\mathbf{A}^{\mathsf{T}}) = \det(\mathbf{A})$. □

8 Friday March 6

Definition 52. A **vector space** is a set V whose elements are called **vectors** and equipped with

(i) an element $\mathbf{0}$ called the **zero vector**,

(ii) a binary operation called **vector addition** which assigns to each pair $(\mathbf{u}, \mathbf{v})$ of vectors another vector $\mathbf{u} + \mathbf{v}$, and

(iii) an operation called **scalar multiplication** which assigns to each number c and each vector $\mathbf{v}$ another vector $c\mathbf{v}$,

such that the following properties hold:

$(\mathbf{u} + \mathbf{v}) + \mathbf{w} = \mathbf{u} + (\mathbf{v} + \mathbf{w})$.	(Additive Associative Law)
$\mathbf{u} + \mathbf{v} = \mathbf{v} + \mathbf{u}$.	(Additive Commutative Law)
$\mathbf{u} + \mathbf{0} = \mathbf{u}$.	(Additive Identity)
$\mathbf{u} + (-1)\mathbf{u} = \mathbf{0}$.	(Additive Inverse)
$c(\mathbf{u} + \mathbf{v}) = c\mathbf{u} + c\mathbf{v}$, $(b + c)\mathbf{u} = b\mathbf{u} + c\mathbf{u}$.	(Distributive Laws)
$(bc)\mathbf{u} = b\big(c\mathbf{u}\big)$.	(Scalar Asscociative Law)
$1\mathbf{u} = \mathbf{u}$.	(Scalar Unit)
$0\mathbf{u} = \mathbf{0}$, $c\mathbf{0} = \mathbf{0}$	(Multiplication by Zero).

Example 53. As noted above in paragraph 18 the set of all $m \times n$ matrices with the operations defined there is a vector space. This vector space is denoted by M_{mn} in the text (see page 272); other textbooks denote it by $\mathbb{R}^{m\times n}$. ($\mathbb{R}$ denotes the set of real numbers.) Note that the text (see page 235 and Remark 50 above) uses $\mathbb{R}^n$ as a synonym for $\mathbb{R}^{n\times 1}$ and has three notations for the elements of $\mathbb{R}^n$:

$$\mathbf{x} = (x_1, x_2, \dots, x_n) = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix}^{\mathsf{T}} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}.$$

Examples 54. Here are some examples of vector spaces.

(i) The set $\mathcal{F}$ of all real valued functions of a real variable.

(ii) The set $\mathcal{P}$ of *all* polynomials with real coefficients.

(iii) The set $\mathcal{P}_n$ of all polynomials of degree $\le n$.

(iv) The set of all solutions of the homogeneous linear differential equation

$$\frac{d^2x}{dt^2} + x = 0.$$

(v) The set of all solutions of *any* homogeneous linear differential equation.

The zero vector in $\mathcal{F}$ is the constant function whose value is zero and the operations of addition and scalar multiplication are defined pointwise, i.e. by

$$(f+g)(x) := f(x) + g(x), \qquad (cf)(x) = cf(x).$$

The set $\mathcal{P}$ is a subspace of $\mathcal{F}$ (a polynomial is a function); in fact, all these vector spaces are subspaces of $\mathcal{F}$. The zero polynomial has zero coefficients, adding two polynomials of degree $\le n$ is the same as adding the coeficients:

$$\begin{aligned}(a_0 + a_1x + \cdots + a_nx^n) + (b_0 + b_1x + \cdots + b_nx^n) &= \\ = (a_0+b_0) + (a_1+b_1)x + \cdots + (a_n+b_n)x^n,\end{aligned}$$

and multiplying a polynomial by number c is the same as multiplying each coefficient by c:

$$c(a_0 + a_1x + \cdots a_nx^n) = ca_0 + ca_1x + \cdots ca_nx^n.$$

Definition 55. A subset $W \subseteq V$ of a vector space V is called a **subspace** iff it is **closed** under the vector space operations, i.e. iff

(i) $\mathbf{0} \in W$,

(ii) $\mathbf{u}, \mathbf{v} \in W \implies \mathbf{u} + \mathbf{w} \in W$, and

(iii) $c \in \mathbb{R}, \mathbf{u} \in W \implies c\mathbf{u} \in W$.

Remark 56. The definition of subspace on page 237 of the text appears *not* to require the condition (i) that $\mathbf{0} \in W$. However that definition *does* specify that W is non empty; this implies that $\mathbf{0} \in W$ as follows. There is an element $\mathbf{u} \in W$ since W is nonempty. Hence $(-1)\mathbf{u} \in W$ by (iii) and $\mathbf{0} = \mathbf{u} + (-1)\mathbf{u} \in W$ by (ii). Conversely, if $\mathbf{0} \in W$, then the set W is nonempty as it contains the element $\mathbf{0}$.

The student is cautioned not to confuse the vector $\mathbf{0}$ with the empty set. The latter is usually denoted by $\emptyset$. The empty set is characterized by the fact that it has no elements, i.e. the statement $x \in \emptyset$ is always false. In particular, $0 \notin \emptyset$. The student should also take care to distinguish the words *subset* and *subspace*. A subspace of a vector space V is a subset of V with certain properties, and not every subset of V is a subspace.

57. A subspace of a vector space is itself a vector space. To decide if a subset W of a vector space V is a subspace you must check that the three properties in Definition 55 hold.

Example 58. The set $\mathcal{P}_n$ of ploynomials of degree $\leq n$ is a subset of the vector space $\mathcal{P}$ (its elements are polynomials) and the set $\mathcal{P}_n$ is a subspace of $\mathcal{P}_m$ if $n \leq m$ (if $n \leq m$ then a p polynomial of degree $\leq n$ has degree $\leq m$). These are also subspace because they are closed under the vector space operations.

9 Monday March 9 – Wednesday March 11

Definition 59. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in a vector space V. The vector $\mathbf{w}$ in V is said to be **linear combination** of the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ iff there exist numbers $x_1, x_2, \ldots, x_k$ such that

$$\mathbf{w} = x_1\mathbf{v}_2 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k.$$

The set of all linear combinations of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is called the **span** of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are said to **span** V iff V is the span of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$, i.e. iff every vector in V is a linear combination of $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$.

Theorem 60. *Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in a vector space V. Then the span $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ is a subspace of V.*

Proof. (See Theorem 1 page 243 of the text.) The theorem says that a linear combination of linear combinations is a linear combination. Here are the details of the proof.

(i) $\mathbf{0}$ is in the span since $\mathbf{0} = 0\mathbf{v}_1 + 0\mathbf{v}_2 + \cdots + 0\mathbf{v}_k$.

(ii) If $\mathbf{v}$ and $\mathbf{w}$ are in the span, there are numbers $a_1, \ldots, b_k$ such that $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$ and $\mathbf{w} = b_1\mathbf{v}_1 + b_2\mathbf{v}_2 + \cdots + b_k\mathbf{v}_k$ so $\mathbf{v} + \mathbf{w} = (a_1 + b_1)\mathbf{v}_1 + (a_2 + b_2)\mathbf{v}_2 + \cdots + (a_k + b_k)\mathbf{v}_k$ so $\mathbf{v} + \mathbf{w}$ is in the span.

(iii) If c is a number and $\mathbf{v}$ is in the span then there are numbers $a_1, \ldots, a_k$ such that $\mathbf{v} = a_1\mathbf{v}_1 + a_2\mathbf{v}_2 + \cdots + a_k\mathbf{v}_k$ so $c\mathbf{v} = ca_1\mathbf{v}_1 + ca_2\mathbf{v}_2 + \cdots + ca_k\mathbf{v}_k$ so $c\mathbf{v}$ is in the span.

Thus we have proved that the span satisfies the three conditions in the definition of subspace so the span is a subspace. □

Example 61. If $\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}$ is an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, then $\mathbf{b}$ is a linear combination of $\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_n$ if and only if the linear system $\mathbf{b} = \mathbf{A}\mathbf{x}$ is consistent, i.e. has a solution $\mathbf{x}$. This is because of the formula

$$\mathbf{A}\mathbf{x} = x_1\mathbf{a}_2 + x_2\mathbf{a}_2 + \cdots + x_n\mathbf{a}_n$$

for $\mathbf{x} = (x_1, x_2, \ldots, x_n)$. The span of the columns $\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n$ is called the **column space** of $\mathbf{A}$.

Definition 62. Let $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ be vectors in a vector space V. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are said to be **independent**[5] iff the only solution of the equation

$$x_1\mathbf{v}_2 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k = 0 \qquad (*)$$

[5]The more precise term **linearly independent** is usually used. We will use the shorter term since this is the only kind of independence we will study in this course.

is the **trivial solution** $x_1 = x_2 = \cdots = x_k = 0$. The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are said to be **dependent** iff they are not independent, i.e. iff there are numbers $x_1, x_2, \ldots, x_k$ not all zero which satisfy $(*)$.

Theorem 63. *The vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are dependent if and only if one of them is in the span of the others.*

Proof. Assume that $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are dependent. Then there are numbers $x_1, x_2 \ldots, x_k$ not all zero such that $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k = \mathbf{0}$. Since the numbers $x_1, x_2, \ldots, x_k$ are not all zero, one of them, say x_i is not zero so

$$\mathbf{v}_i = -\frac{x_1}{x_i}\mathbf{v}_1 - \cdots - \frac{x_{i-1}}{x_i}\mathbf{v}_{i-1} - \frac{x_{i+1}}{x_i}\mathbf{v}_{i+1} - \cdots - \frac{x_k}{x_i}\mathbf{v}_k,$$

i.e. $\mathbf{v}_i$ is a linear combination of $\mathbf{v}_1, \ldots.\mathbf{v}_{i-1}, \mathbf{v}_{i+1} \ldots, \mathbf{v}_k$. Suppose conversely that $\mathbf{v}_i$ is a linear combination of $\mathbf{v}_1, \ldots.\mathbf{v}_{i-1}, \mathbf{v}_{i+1} \ldots, \mathbf{v}_k$. Then there are numbers $c_1, \ldots, c_{i-1}, c_{i+1}, \ldots, c_k$ such that $\mathbf{v}_i = c_1\mathbf{v}_1 + \cdots + c_{i-1}\mathbf{v}_{i-1} + c_{i+1}\mathbf{v}_{i+1} + \cdots + c_k\mathbf{v}_k$. Then $x_1\mathbf{v}_1 + x_2\mathbf{v}_2 + \cdots + x_k\mathbf{v}_k = \mathbf{0}$ where $x_j = c_j$ for $j \neq i$ and $x_i = -1$. Since $-1 \neq 0$ the numbers $x_1, x_2, \ldots, x_k$ are not all zero and so the vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$ are dependent. □

Remark 64. (A pedantic quibble.) The text says things like "the set of vectors $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k$" is independent" but it is better to use the word "sequence" instead of "set". The sets $\{\mathbf{v}, \mathbf{v}\}$ and $\{\mathbf{v}\}$ are the same (both consist of the single element $\mathbf{v}$) but if $\mathbf{v} \neq \mathbf{0}$ the sequence whose one and only element is $\mathbf{v}$ is independent (since $c\mathbf{v} = \mathbf{0}$ only if $c = 0$) whereas the two element sequence $\mathbf{v}, \mathbf{v}$ (same vector repeated) is always dependent since $c_1\mathbf{v} + c_2\mathbf{v} = \mathbf{0}$ if $c_1 = 1$ and $c_2 = -1$.

Definition 65. A **basis** for a vector space V is a sequence $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ of vectors in V which both spans V and is independent.

Theorem 66. *If $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$ is a basis for V and $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_m$ is a basis for V, then $m = n$.*

This is Theorem 2 on page 251 of the text. We will prove it next time. It justifies the following

Definition 67. The **dimension** of a vector space is the number of elements in some (and hence every) basis.

Remark 68. It can happen that there are arbitrarily long independent sequences in V. For example, this is the case if $V = \mathcal{P}$, the space of all polynomials: for every n the vectors $1, x, x^2, \ldots, x^n$ are independent. In this case we say that V is **infinite dimensional**.

10 Friday March 13

Proof of Theorem 66. Let $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ be two sequences of vectors in a vector space V. It is enough to prove

(†) If $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ span V, and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are independent, then $n \le m$.

To deduce Theorem 66 from this we argue as follows: If both sequences $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ and $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are bases then the former spans and the latter is independent so $n \le m$. Reversing the roles gives $m \le n$. If $n \le m$ and $m \le n$, then $m = n$. To prove the assertion (†) is enough to prove the contrapositive:

> If $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ span V and $n > m$, then $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are dependent.

To prove the contrapositive note that because $\mathbf{w}_1, \mathbf{w}_2, \dots, \mathbf{w}_m$ span there are (for each $j = 1, \dots, n$) constants $a_{1j}, \dots, a_{mj}$ such that

$$\mathbf{v}_j = \sum_{i=1}^{m} a_{ij} \mathbf{w}_i.$$

This implies that for any numbers $x_1, x_2, \dots, x_n$ we have

$$\sum_{j=1}^{n} x_j \mathbf{v}_j = \sum_{j=1}^{n} x_j \left(\sum_{i=1}^{m} a_{ij} \mathbf{w}_i \right) = \sum_{i=1}^{m} \left(\sum_{j=1}^{n} a_{ij} x_j \right) \mathbf{w}_i. \qquad (\#)$$

Since $n > m$ the homogeneous linear system

$$\sum_{j=1}^{n} a_{ij} x_j = 0, \qquad i = 1, 2, \dots, m \qquad (\flat)$$

has more unknowns than equations so there is a nontrivial solution $\mathbf{x} = (x_1, x_2, \dots, x_n)$. The left hand side of ($\flat$) is the coefficient of $\mathbf{w}_i$ in (#) so ($\flat$) implies that $\sum_{j=1}^{n} x_j \mathbf{v}_j = \mathbf{0}$, i.e. that $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_n$ are dependent. □

Definition 69. For an $m \times n$ matrix $\mathbf{A}$

(i) The **row space** is the span of the rows of $\mathbf{A}$.

(ii) The **column space** is the span of the columns of $\mathbf{A}$.

(iii) The **null space** is the set of all solutions $\mathbf{x} \in \mathbb{R}^n$ of the homogeneous system $\mathbf{Ax} = \mathbf{0}$.

The dimension of the row space of $\mathbf{A}$ is called the **rank** of $\mathbf{A}$. The text calls the dimension of the row space the **row rank** and the dimension of the columns space the **column rank** but Theorem 72 below says that these are equal.

Theorem 70 (Equivalent matrices)**.** *Suppose that* $\mathbf{A}$ *and* $\mathbf{B}$ *are equivalent* $m \times n$ *matrices. Then*

(i) $\mathbf{A}$ *and* $\mathbf{B}$ *have the same null space.*

(ii) $\mathbf{A}$ *and* $\mathbf{B}$ *have the same row space.*

Proof. Assume that $\mathbf{A}$ and $\mathbf{B}$ are equivalent. Then $\mathbf{B} = \mathbf{MA}$ where $\mathbf{M} = \mathbf{E}_1\mathbf{E}_2 \cdots \mathbf{E}_k$ is a product of elementary matrices. If $\mathbf{Ax} = \mathbf{0}$ then $\mathbf{Bx} = \mathbf{MAx} = \mathbf{M0} = \mathbf{0}$. Similarly if $\mathbf{Bx} = \mathbf{0}$ then $\mathbf{Ax} = \mathbf{M}^{-1}\mathbf{Bx} = \mathbf{M}^{-1}\mathbf{0} = \mathbf{0}$. Hence $\mathbf{Ax} = \mathbf{0} \iff \mathbf{Bx} = \mathbf{0}$ which shows that $\mathbf{A}$ and $\mathbf{B}$ have the same null space. Another way to look at it is that performing an elementary row operation doesn't change the space of solutions of the corresponding homogeneous linear system. This proves (i)

Similarly performing an elementary row operation doesn't change the row space. This is because if $\mathbf{E}$ is an elementary matrix then each row of $\mathbf{EA}$ is either a row of $\mathbf{A}$ or is a linear combination of two rows of $\mathbf{A}$ so a linear combination of rows of $\mathbf{EA}$ is also a linear combination of rows of $\mathbf{A}$ (and vice versa since $\mathbf{E}^{-1}$ is also an elementary matrix). This proves (ii). □

Theorem 71. *The rank of a matrix* $\mathbf{A}$ *is the number* r *of non zero rows in the reduced echelon form of* $\mathbf{A}$*.*

Proof. By part (ii) of Theorem 70 it is enough to prove this for a matrix which is in reduced echelon form. The non zero rows clearly span the row space (by the definition of the row space) and they are independent since the identity matrix appears as an $r \times r$ submatrix. □

Theorem 72. *The null space of an* $m \times n$ *matrix has dimension* $n - r$ *where* r *is the rank of the matrix.*

Proof. The algorithm on page 254 of the text finds a basis for the null space. You put the matrix in reduced echelon form. The number of leading variables is r so there are $n - r$ free variables. A basis consists of the solutions of the system obtained by setting one of the free variables to one and the others to zero. $\square$

Spring recess. Mar 14-22 (S-N)

11 Monday March 23

73. Theorem 70 says that equivalent matrices have the same row space, but they need not have the same column space. The matrices

$$\mathbf{A} = \begin{bmatrix} 1 & 0 \\ 1 & 0 \end{bmatrix} \qquad \text{and} \qquad \mathbf{B} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

are equivalent and the row space of each is the set of multiples of the row $\begin{bmatrix} 1 & 0 \end{bmatrix}$, but the column spaces are different: the column space of $\mathbf{A}$ consists of all multiples of the column $(1, 1)$ while the column space of $\mathbf{B}$ consists of all multiples of column $(1, 0)$. However

Theorem 74. *The row rank equals the column rank, i.e. the column space and row space of an $m \times n$ matrix $\mathbf{A}$ have the same dimension.*

Proof. Theorem 63 says that if a sequence $\mathbf{v}_1, \ldots, \mathbf{v}_n$ of vectors is dependent then one of them is a linear combination of the others. This vector can be deleted without changing the span. In particular, if the columns of a matrix are dependent we can delete one of them without changing the column space. This process can be repeated until the vectors that remain are independent. The remaining vectors then form a basis. Thus a basis for the column space of $\mathbf{A}$ can be selected from the columns of $\mathbf{A}$. The algorithm in the text on page 259 tells us that these can be the **pivot columns** of $\mathbf{A}$: these are the columns corresponding to the leading variables in the reduced echelon form.

Let $\mathbf{a}_k$ be the kth column of $\mathbf{A}$ and $\mathbf{r}_k$ be the kth column of the redced echelon form $\mathbf{R}$ of $\mathbf{A}$. Then

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1 & \mathbf{a}_2 & \cdots & \mathbf{a}_n \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{bmatrix},$$

and $\mathbf{MA} = \mathbf{R}$ where $\mathbf{M}$ is the invertible matrix which is the product of the elementary matrices used to transform $\mathbf{A}$ to its reduced echelon form $\mathbf{R}$. Now matrix multiplication distributes over concatenation:

$$\mathbf{MA} = \begin{bmatrix} \mathbf{Ma}_1 & \mathbf{Ma}_2 & \cdots & \mathbf{Ma}_n \end{bmatrix} = \mathbf{R} = \begin{bmatrix} \mathbf{r}_1 & \mathbf{r}_2 & \cdots & \mathbf{r}_n \end{bmatrix},$$

so

$$\mathbf{Ma}_k = \mathbf{r}_k, \qquad \text{and} \qquad \mathbf{r}_k = \mathbf{M}^{-1}\mathbf{a}_k$$

for $k = 1, 2, \ldots, n$. After rearranging the columns of $\mathbf{R}$ and rearranging the columns of $\mathbf{A}$ the same way we may assume that the first r columns of $\mathbf{R}$ are the first r columns $\mathbf{e}_1, \mathbf{e}_2, \ldots, \mathbf{e}_r$ of the identity matrix and the last $n - r$ rows of $\mathbf{R}$ are zero. Then each of the last $n - r$ columns of $\mathbf{R}$ is a linear combination of the first r columns so (multiplying by $\mathbf{M}$) each of the last $n - r$ columns of $\mathbf{A}$ is a linear combination of the first r columns (with the same coefficients). Hence the first columns of $\mathbf{A}$ span the column space of $\mathbf{A}$. If some linear combination of the first r columns of $\mathbf{A}$ is zero, then (multiplying by $\mathbf{M}^{-1}$) the same linear combination of the first r columns is zero. But the first r columns of $\mathbf{R}$ are the first r columns of the identity matrix so the coefficients must be zero. Hence the first r columns of $\mathbf{A}$ are independent. □

Example 75. The following matrices were computer generated.

$$\mathbf{A} = \begin{bmatrix} 1 & 3 & 19 & 23 \\ -1 & -2 & -14 & -17 \\ -2 & -6 & -38 & -46 \\ -2 & -7 & -43 & -52 \end{bmatrix}, \qquad \mathbf{R} = \begin{bmatrix} 1 & 0 & 4 & 5 \\ 0 & 1 & 5 & 6 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\mathbf{M} = \begin{bmatrix} -2 & -1 & -3 & 2 \\ -6 & 0 & -2 & -1 \\ 5 & 3 & -2 & 3 \\ 1 & 1 & -1 & 1 \end{bmatrix}, \qquad \mathbf{M}^{-1} = \begin{bmatrix} 1 & 3 & 10 & -29 \\ -1 & -2 & -7 & 21 \\ -2 & -6 & -19 & 55 \\ -2 & -7 & -22 & 64 \end{bmatrix}.$$

The matrix $\mathbf{R}$ is the reduced echelon form of $\mathbf{A}$ and $\mathbf{MA} = \mathbf{R}$. The pivot columns are the first two columns. The third column of $\mathbf{R}$ is $4\mathbf{e}_1 + 5\mathbf{e}_2$ and the third column of $\mathbf{A}$ is $\mathbf{a}_3 = 4\mathbf{a}_1 + 5\mathbf{a}_2$. The fourth column of $\mathbf{R}$ is $5\mathbf{e}_1 + 6\mathbf{e}_2$ and the fourth column of $\mathbf{A}$ is $\mathbf{a}_4 = 5\mathbf{a}_1 + 6\mathbf{a}_2$. The first two columns of $\mathbf{A}$ are the same as the first two columns of $\mathbf{M}^{-1}$.

11.1 Wednesday March 22

The following material is treated in Section 4.6 of the text. We may not have time to cover it in class so you should learn it on your own,

Definition 76. The **inner product** of two vectors $\mathbf{u} = (u_1, u_2, \ldots, u_n)$ and $\mathbf{v} = (v_1, v_2, \ldots, v_n)$ in $\mathbb{R}^n$ is denoted $\langle \mathbf{u}, \mathbf{v} \rangle$ and defined by

$$\langle \mathbf{u}, \mathbf{v} \rangle = u_1 v_2 + u_2 v_2 + \cdots + u_n v_n.$$

It was called the **dot product** in Math 222. It can also be expressed in terms of the transpose operation as

$$\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^{\mathsf{T}} \mathbf{v}.$$

The **length** $|\mathbf{u}|$ of the vector $\mathbf{u}$ is defined as

$$|\mathbf{u}| := \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}.$$

Two vectors are called **orthogonal** iff their inner product is zero.

77. The inner product satisfies the following.

(i) $\langle \mathbf{u}, \mathbf{v} \rangle = \langle \mathbf{v}, \mathbf{u} \rangle$.

(ii) $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.

(iii) $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$.

(iv) $\langle \mathbf{u}, \mathbf{u} \rangle \geq 0$ and $\langle \mathbf{u}, \mathbf{u} \rangle = 0 \iff \mathbf{u} = 0$.

(v) $|\langle \mathbf{u}, \mathbf{v} \rangle| \leq |\mathbf{u}|\,|\mathbf{v}|$.

(vi) $|\mathbf{u} + \mathbf{v}| \leq |\mathbf{u}| + |\mathbf{v}|$.

The inequality (v) is called the **Cauchy Schwartz Inequality**. It justifies defining the **angle** θ between two nonzero vectors $\mathbf{u}$ and $\mathbf{v}$ by the formula

$$\langle \mathbf{u}, \mathbf{v} \rangle = |\mathbf{u}|\,|\mathbf{v}|\,\cos\theta$$

Thus two vectors are orthogonal iff the angle between them is $\pi/2$. The inequality (vi) is called the **triangle inequality**.

Theorem 78. *Suppose that the vectors* $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ *are non zero and pairwise orthogonal, i.e.* $\langle \mathbf{v}_i, \mathbf{v}_j \rangle = \mathbf{0}$ *for* $i \neq j$*. Then the sequence* $\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_k$ *is independent.*

Definition 79. Let V be a subspace of $\mathbb{R}^n$. The **orthogonal complement** is the set $V^\perp$ of vectors which are orthogonal to all the vectors in V, in other words

$$\mathbf{w} \in V^\perp \iff \langle v, w \rangle = 0 \text{ for all } \mathbf{v} \in V.$$

Theorem 80. *The column space of* $\mathbf{A}^\mathsf{T}$ *is the orthogonal complement to the null space of* $\mathbf{A}$*.*

Exam II. Friday Mar 27