MATH 20610, LINEAR ALGEBRA
NOTATION, TERMINOLOGY AND BASIC FACTS

JEFFREY DILLER

ABSTRACT. The following is intended as a helpful summary and some further comments concerning the linear algebra we've been learning. It's meant to complement and reinforce lecture and homework but not certainly not replace them. If you see anything suspicious, ask about it—there are likely some typos. Things highlighted in red are important enough that I expect you to be able to define/state them yourself on quizzes and exams.

1. Matrices and vectors

Definition 1.1. A $m \times n$ *matrix* is a rectangular array of numbers

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{pmatrix}$$

The number a_{ij} is called the ij-entry of A.

Definition 1.2. A *column n-vector* is an $n \times 1$ matrix

$$\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}$$

A *row n-vector* is a $1 \times n$ matrix

$$\mathbf{w} = \begin{pmatrix} w_1 & \dots & w_n \end{pmatrix}.$$

More often than not our vectors will be column vectors, and if the context is clear enough, we'll just say 'vector' instead of 'column n-vector' and not make much distinction between row and column vectors. We'll use $\mathbb{R}^n$ to denote the set of all n-vectors.

Definition 1.3. Let $A = (a_{ij})$ and $B = (b_{ij})$ be $m \times n$ matrices and $c \in \mathbb{R}$ be a scalar. Then we let

- $A + B$ denote the matrix whose ij-entry is $a_{ij} + b_{ij}$, and
- cA denote the matrix whose ij-entry is ca_{ij}.

One can check that matrix arithmetic follows many of the usual rules: i.e. if A, B and C are $m \times n$ matrices and $a, b \in \mathbb{R}$ are scalars, then

- $A + B = B + A$;
- $(A + B) + C = A + (B + C)$;
- $(a + b)C = aC + bC$; and
- $a(B + C) = aB + aC$.

Date: December 6, 2025.

Finally, if $0_{m\times n}$ denotes the $m \times n$ matrix whose entries are all zeros, then

- $a0_{m\times n} = 0_{m\times n}$; and
- $0_{m\times n} + A = A$.

Note that we use $\mathbf{0}_n \in \mathbb{R}^n$ to denote the *vector* whose entries are all zero. And we will often drop the subscripts, writing 0 for the zero matrix and $\mathbf{0}$ for the zero vector when there isn't much danger of confusion.

The terminology in the following definition is not used in Bretscher, but I will use it freely.

Definition 1.4. Let A be a matrix.

- The leftmost non-zero entry in a given row of a matrix is called the *pivot* for that row.
- A is said to be in *(row) echelon form* if the pivot entry (if any) in each row lies to the left of the pivot entry (if any) in the next row.
- If A is in echelon form, then a *pivot column* of A is one containing the pivot from some row of A.

Definition 1.5. A matrix A is said to be in *reduced (row) echelon form* (RREF for short) if it is in echelon form and, additionally,

- all pivot entries are equal to 1; and
- all entries in each pivot column, except for the pivots, are equal to 0.

Definition 1.6. Matrices A and B are said to be *row equivalent* if one can be transformed into the other by a sequence of elementary row operations.

Theorem 1.7. *Every matrix is row equivalent to a matrix in reduced echelon form.*

In fact, each matrix is row equivalent to *exactly one* matrix in reduced echelon form, but this fact is a little difficult to justify.

Definition 1.8. [1] The *rank* of a matrix A is the number of pivots in the RREF matrix that is row equivalent to A.

Definition 1.9. If $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are vectors and $c_1, \dots c_k \in \mathbb{R}$ are scalars, then we call the vector

$$c_1\mathbf{v}_1 + \dots c_k\mathbf{v}_k$$

a *linear combination* of $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Definition 1.10. The *dot product* of two vectors $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix}, \mathbf{w} = \begin{pmatrix} w_1 \\ \vdots \\ w_n \end{pmatrix} \in \mathbb{R}^n$ is the scalar quantity

$$\mathbf{v} \cdot \mathbf{w} := v_1w_1 + \dots v_nw_n.$$

Again, the dot product shares many properties with standard multiplication of real numbers. If $\mathbf{uv}, \mathbf{w} \in \mathbb{R}^n$ are vectors and $a, b \in \mathbb{R}$ are scalars, then

- $\mathbf{v} \cdot \mathbf{w} = \mathbf{w} \cdot \mathbf{v}$;
- $\mathbf{u} \cdot (\mathbf{v} + \mathbf{w}) = \mathbf{u} \cdot \mathbf{v} + \mathbf{u} \cdot \mathbf{w}$;
- $\mathbf{0} \cdot \mathbf{v} = 0$; and

[1] This definition is superseded by Definition 6.18 below.

- $a(\mathbf{v} \cdot \mathbf{w}) = (a\mathbf{v}) \cdot \mathbf{w} = \mathbf{v} \cdot (a\mathbf{w})$.

Definition 1.11. Let A be an $m \times n$ matrix with columns $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^m$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ be another vector. Then the *product* of A and $\mathbf{x}$ is the vector

$$A\mathbf{x} = x_1\mathbf{v}_1 + \ldots x_n\mathbf{v}_n \in \mathbb{R}.$$

Alternatively, if $\mathbf{w}_1, \ldots, \mathbf{w}_m \in \mathbb{R}^n$ are the *rows* of A, then

$$A\mathbf{x} = \begin{pmatrix} \mathbf{w}_1 \cdot \mathbf{x} \\ \vdots \\ \mathbf{w}_m \cdot \mathbf{x} \end{pmatrix}.$$

Proposition 1.12. *Let A and B be $m \times n$ matrices, $\mathbf{x}, \mathbf{y} \in \mathbb{R}^n$ be vectors and $c \in \mathbb{R}$ be a scalar. Then*

- $A(\mathbf{x} + \mathbf{y}) = A\mathbf{x} + A\mathbf{y}$*;*
- $c(A\mathbf{x}) = (cA)\mathbf{x} = A(c\mathbf{x})$*;*
- $(A + B)\mathbf{x} = A\mathbf{x} + B\mathbf{x}$*.*

The $n \times n$ identity matrix is

$$I_{n\times n} := \begin{pmatrix} 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

That is, I is a 'square' matrix, with the same number of rows as columns, and its ij-entry is 1 if $i = j$ and 0 if $i \neq j$.

Proposition 1.13. *For any vector $\mathbf{v} \in \mathbb{R}^n$, we have $0_{m\times n}\mathbf{v} = \mathbf{0}_m$ and $I_{n\times n}\mathbf{v} = \mathbf{v}$.*

2. Linear systems

Definition 2.1. An $m \times n$ *linear system* is a list of equations

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

Here $a_{ij} \in \mathbb{R}$ are the *coefficients* of the system, $b_1, \dots, b_m \in \mathbb{R}$ are given numbers and $x_1, \dots, x_n \in \mathbb{R}$ are the variables (i.e. unknown numbers to be solved for).

There are three possibilities for a given linear system:

- The system has no solutions $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$;
- The system has exactly one solution $\mathbf{x} \in \mathbb{R}^n$;
- The system has infinitely many solutions $\mathbf{x} \in \mathbb{R}^n$.

In the first case, we call the system *inconsistent*. In the other two cases, we call it *consistent*.

There are several different and useful ways to express a linear system. First of all, one can write it as $m \times (n+1)$ *augmented matrix*

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} & | & b_1 \\ a_{21} & a_{22} & \dots & a_{2n} & | & b_2 \\ \vdots & \vdots & & \vdots & \vdots & \\ a_{m1} & a_{m2} & \dots & a_{mn} & | & b_m. \end{bmatrix}$$

Using Gauss-Jordan elimination, one can perform elementary row operations on the augmented matrix to put it in reduced echelon form. If, at that point, there is a pivot in the very last column, the linear system is inconsistent. If every column *except* the last one has a pivot, then the system has exactly one solution. If neither of these are true (i.e. there is no pivot in the last column and no pivot in at least one of the other columns), then the system has infinitely many solutions.

We call the $m \times n$ matrix $A = (a_{ij})$ the *coefficient matrix* of the system. Setting $\mathbf{b} := \begin{pmatrix} b_1 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^m$, we can rewrite the augmented matrix very compactly as

$$\begin{bmatrix} A & | & \mathbf{b} \end{bmatrix}.$$

If $\mathbf{w}_1, \dots, \mathbf{w}_m \in \mathbb{R}^m$ are the rows of A, and we set $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ then we can write the linear system as

$$\begin{array}{ccc} \mathbf{w}_1 \cdot \mathbf{x} & = & b_1 \\ & \vdots & \\ \mathbf{w}_m \cdot \mathbf{x} & = & b_m. \end{array}$$

In terms of the columns $\mathbf{v}_1, \ldots, \mathbf{w}_n \in \mathbb{R}^m$ of A, the system becomes a vector equation

$$x_1\mathbf{v}_1 + \cdots + x_n\mathbf{v}_n = \mathbf{b}.$$

Finally, this can be rewritten as a matrix/vector equation

$$A\mathbf{x} = \mathbf{b},$$

which is probably the best way of all to write a linear system.

3. Linear Transformations

Definition 3.1. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a *linear transformation* if for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$ we have

- $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$, and
- $T(c\mathbf{v}) = cT(\mathbf{v})$.

The two conditions in this definition are often rephrased by saying that 'T commutes with vector addition and scalar multiplication.' Together they imply that 'T commutes with linear combinations':

Proposition 3.2. *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, and $c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in \mathbb{R}^n$ is a linear combination, then*

$$T(c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k) = c_1T(\mathbf{v}_1) + \cdots + c_kT(\mathbf{v}_k).$$

The prototypical, and in some sense *only*, example of a linear transformation is a *matrix transformation*: if A is an $m \times n$ matrix, they we get a function $T = T_A : \mathbb{R}^n \to \mathbb{R}^m$ given by $T(\mathbf{x}) = A\mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. The next theorem explains the relationship between linear and matrix transformation. In order to state it we introduce some new notation.

Definition 3.3. Given $j \in \{1, \ldots, n\}$, the jth *standard basis vector* $\mathbf{e}_j \in \mathbb{R}^n$ is the vector jth entry is 1 and whose other entries all equal 0.

The point is that any other vector $\mathbf{x} = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix} \in \mathbb{R}^n$ is easily written as a linear combination

$$\mathbf{x} = x_1\mathbf{e}_1 + \cdots + x_n\mathbf{e}_n$$

of standard basis vectors.

Theorem 3.4. *A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation if and only if it is a matrix transformation. In this case, the matrix $[T]$ for T is given column-wise by*

$$[T] = \begin{bmatrix} T(\mathbf{e}_1) & \ldots & T(\mathbf{e}_n) \end{bmatrix},$$

where $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ are the standard basis vectors.

Example 3.5. Let $\mathrm{id} : \mathbb{R}^n \to \mathbb{R}^n$ denote the *identity transformation*, given by $\mathrm{id}(\mathbf{x}) = \mathbf{x}$ for all $\mathbf{x} \in \mathbb{R}^n$. Then the matrix for id is the identity matrix, i.e.

$$[\mathrm{id}] = I_{n\times n}.$$

Also, if $0 : \mathbb{R}^n \to \mathbb{R}^n$ denotes the *zero transformation*, given by $0(\mathbf{x}) = \mathbf{0}$ for all $\mathbf{x} \in \mathbb{R}^n$, then the matrix for 0 is the $n \times n$ 0-matrix, i.e.

$$[0] = 0_{n\times n}$$

4. Geometry and linear transformations

In this section we describe matrices associated to a variety of geometrically defined transformations. Our first result concerns plane rotations and is really more of a definition than a proposition.

Proposition 4.1. *Given $\theta \in \mathbb{R}$, Let $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ be the function defined by setting $T(\mathbf{v})$ equal to the vector obtained by rotating $\mathbf{v}$ counterclockwise by θ. Then R_θ is a linear transformation, and its matrix is*

$$[R_\theta] = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

Our other geometric transformations require a little warmup.

Definition 4.2. The *length* (or *norm*) of a vector $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ is the quantity

$$\|\mathbf{v}\| := \sqrt{\mathbf{v}\cdot\mathbf{v}} = \sqrt{v_1^2 + \cdots + v_n^2} \in \mathbb{R}.$$

The three key properties of length are the following

Proposition 4.3. *For any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$, we have*

- *(positivity) $\|\mathbf{v}\| \geq 0$, and $\|\mathbf{v}\| = 0$ if and only if $\mathbf{v} = \mathbf{0}$;*
- *(homogeneity) $\|c\mathbf{v}\| = |c|\ \|\mathbf{v}\|$;*
- *(triangle inequality) $\|\mathbf{v}+\mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$.*

The first two of these properties are fairly straightforward to check. The third is harder, and I won't really prove it til later in the course.

Definition 4.4. If $\mathbf{u} \in \mathbb{R}^n$ has length $\|\mathbf{u}\| = 1$, then we call $\mathbf{u}$ a *unit* vector.

If $\mathbf{v} \in \mathbb{R}^n$ is a non-zero vector, then one checks that $\frac{\mathbf{v}}{\|\mathbf{v}\|}$ is unit vector.

Pythagorus' Theorem, together with the relationship between length and dot product, is the key to the following definition.

Definition 4.5. Vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are called *orthogonal* (or *perpendicular*) if $\mathbf{v}\cdot\mathbf{w} = 0$.

Theorem 4.6 (Orthogonal decomposition)**.** *Let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector. Then any other vector $\mathbf{w} \in \mathbb{R}^n$ can be written in exactly one way as a sum*

$$\mathbf{w} = \mathbf{w}_\parallel + \mathbf{w}_\perp,$$

where $\mathbf{w}_\parallel$ is a scalar multiple of $\mathbf{v}$ and and $\mathbf{w}_\perp$ is orthogonal to $\mathbf{v}$. The vector $\mathbf{w}_\parallel$ is given by

$$\mathbf{w}_\parallel = \frac{\mathbf{w}\cdot\mathbf{v}}{\mathbf{v}\cdot\mathbf{v}}\mathbf{v}.$$

Definition 4.7. Let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector. Then the set $L \subset \mathbb{R}^n$ of all scalar multiples of $\mathbf{v}$ is the *line through* $\mathbf{0}$ *and* $\mathbf{v}$.

Definition 4.8. Let $\mathbf{v} \in \mathbb{R}^n$ be a non-zero vector and $L \subset \mathbb{R}^n$ be the line through $\mathbf{0}$ and $\mathbf{v}$. Then the vector $\mathrm{proj}_{\mathbf{v}}(\mathbf{w}) := \mathbf{w}_\parallel$ in Theorem 4.6 is called the *orthogonal projection of* $\mathbf{w}$ *onto L (or onto $\mathbf{v}$).*

Proposition 4.9. *Let* $\mathbf{v} = \begin{pmatrix} v_1 \\ \vdots \\ v_n \end{pmatrix} \in \mathbb{R}^n$ *be a non-zero vector. Then* $\mathrm{proj}_L : \mathbb{R}^n \to \mathbb{R}^n$ *is a linear transformation, and its matrix is given (column-wise) by the formula*

$$[\mathrm{proj}_L] = \frac{1}{\|\mathbf{v}\|^2} \begin{bmatrix} v_1\mathbf{v} & \dots v_n\mathbf{v} \end{bmatrix}.$$

Proposition 4.10. *Let* $\mathbf{v} \in \mathbb{R}^n$ *be a non-zero vector and* $L \subset \mathbb{R}^n$ *denote the line joining* $\mathbf{0}$ *to* $\mathbf{v}$. *Let* $\mathrm{ref}_L : \mathbb{R}^n \to \mathbb{R}^n$ *be the function defined by setting* $\mathrm{ref}_L(\mathbf{w})$ *equal to the reflection of* $\mathbf{w}$ *through* $\mathbf{v}$. *Then* ref_L *is a linear transformation given by*

$$\mathrm{ref}_L(\mathbf{w}) = 2\,\mathrm{proj}_L(\mathbf{w}) - \mathbf{w}.$$

It's matrix is therefore given by

$$[\mathrm{ref}_L] = 2[\mathrm{proj}_L] - I$$

5. Matrix Multiplication and Composition of Linear Transformations

Definition 5.1. Let $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$ be a $p \times n$ matrix and B be an $m \times p$ matrix. Then the *matrix product* of B and A is the $m \times n$ matrix

$$BA := \begin{bmatrix} B\mathbf{v}_1 & \dots & B\mathbf{v}_n \end{bmatrix}.$$

Note that it's important in this definition that the number of columns of B matches the number of rows of A. The main reason for all this is the following.

Theorem 5.2. *Let $S : \mathbb{R}^n \to \mathbb{R}^p$ and $R : \mathbb{R}^p \to \mathbb{R}^m$ be linear transformations. Then $R \circ S : \mathbb{R}^n \to \mathbb{R}^m$ is also a linear transformation, and its matrix is given by the formula*

$$[R \circ S] = [R][S].$$

Theorem 5.3. *Let A, B, C be matrices and $\lambda \in \mathbb{R}$ be a scalar. Then*

- $AI = IA = A$;
- $\lambda(AB) = (\lambda A)B = A(\lambda B)$;
- $A(B + C) = AB + AC$;
- $(A + B)C = AC + BC$;
- $A(BC) = (AB)C$;

provided the sizes of the matrices involved are compatible with each other.

5.1. Matrix inverses.

Definition 5.4. A matrix A is *invertible* if there is another matrix B such that $AB = BA = I$. We then say that B is the *inverse* of A and write $A^{-1} = B$.

Proposition 5.5. *A matrix A has at most one inverse.*

Theorem 5.6. *If an $m \times n$ matrix A is invertible, then for any $\mathbf{b} \in \mathbb{R}^m$, the linear system $A\mathbf{x} = \mathbf{b}$ has exactly one solution.*

That $A\mathbf{x} = \mathbf{b}$ has at most one solution means that A is row equivalent to a RREF matrix with a pivot in every column. Then $A\mathbf{x} = \mathbf{b}$ has at least one solution for any $\mathbf{b}$ means that A is row equivalent to an RREF matrix with a pivot in every row. Since no row or column of an RREF matrix contains more than one pivot, the condition that $A\mathbf{x} = \mathbf{b}$ has exactly one solution for all vectors $\mathbf{b} \in \mathbb{R}^m$ implies that A has the same number of rows as columns.

Corollary 5.7. *If A is invertible, then A is square.*

Even square matrices needn't be invertible though. The next result gives several different ways to tell the difference between invertible and non-invertible square matrices.

Corollary 5.8. *The following are equivalent for an $n \times n$ matrix A.*

1. *A is invertible.*
2. *A has rank n.*
3. *A is row equivalent to the $n \times n$ identity matrix.*
4. *There is an $n \times n$ matrix B such that $AB = I$.*
5. *There is an $n \times n$ matrix B such that $BA = I$.*

In the case of 2×2 matrices, there is a reasonably easy to remember formula for A^{-1}.

Proposition 5.9. *A 2×2 matrix $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ is invertible if and only if $ad - bc \neq 0$. The inverse of A is then given by*

$$A^{-1} = \frac{1}{ad - bc} \begin{pmatrix} d & -b \\ -c & a \end{pmatrix}.$$

For non-zero real numbers a, b we know that $\frac{1}{ab} = \frac{1}{a} \cdot \frac{1}{b}$. The analogous fact for matrices is the following.

Proposition 5.10. *If A and B are invertible $n \times n$ matrices, then*

- *A^{-1} is invertible, with $(A^{-1})^{-1} = A$;*
- *A^n is invertible for any positive integer n, with $(A^n)^{-1} = A^{-n}$;*
- *AB is also invertible, with $(AB)^{-1} = B^{-1}A^{-1}$.*

Recall that the reason we defined matrix multiplication the way we did is that matrix multiplication corresponds to composition of linear transformations. So maybe it's not surprising that invertible matrices correspond to invertible linear transformations.

Definition 5.11. A function $T : \mathbb{R}^n \to \mathbb{R}^m$ is *invertible* if there is a function $S : \mathbb{R}^m \to \mathbb{R}^n$ such that $S \circ T = \mathrm{id}$ and $T \circ S = \mathrm{id}$. We then write $T^{-1} = S$.

Theorem 5.12. *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a invertible linear transformation, then T^{-1} is also linear. Hence $m = n$ and the matrices for the two transformations satisfy $[T^{-1}] = [T]^{-1}$.*

6. Span, Linear Independence, Subspace and Basis

Definition 6.1. The *span* of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k \in \mathbb{R}^n$ is the set

$$\operatorname{span}\{\mathbf{v}_1, \ldots, \mathbf{v}_k\} := \{c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k \in \mathbb{R}^n : c_1, \ldots, c_k \in \mathbb{R}\}$$

of all linear combinations of $\mathbf{v}_1, \ldots, \mathbf{v}_k$.

Some examples:

- $\operatorname{span}(\mathbf{0}) = \{\mathbf{0}\}$
- $\operatorname{span}(\mathbf{e}_1, \ldots, \mathbf{e}_n) = \mathbb{R}^n$
- if $\mathbf{v} \in \mathbb{R}^n$ is non-zero, then $\operatorname{span}(\mathbf{v})$ is a *line* through $\mathbf{0}$;
- if $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are non-zero and not multiples of each other, then $\operatorname{span}(\mathbf{v}, \mathbf{w})$ is a plane through $\mathbf{0}$.

Here are some other sets that turn out to be spans.

Definition 6.2. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation.

- The *image* (alternatively, *range*) of T is the set

$$\operatorname{image}(T) := \{T(\mathbf{x}) \in \mathbb{R}^m : \mathbf{x} \in \mathbb{R}^n.$$

- The *kernel* of T is the set

$$\ker(T) := \{\mathbf{x} \in \mathbb{R}^n : T(\mathbf{x}) = \mathbf{0}\}$$

The following fact follows more or less directly from definitions.

Proposition 6.3. *Suppose $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation with matrix $[T] = A$. Then $\operatorname{image}(T)$ is the span of the columns of A and $\ker(T)$ is the set of solutions $\mathbf{x} \in \mathbb{R}^n$ of $A\mathbf{x} = \mathbf{0}$.*

We remark that a linear system of the form $A\mathbf{x} = \mathbf{0}$ is called *homogeneous*, whereas a linear system $A\mathbf{x} = \mathbf{b}$ for $\mathbf{b} \neq \mathbf{0}$ is called *inhomogeneous*.

Hence, following Bretscher, we will also write $\operatorname{image}(A)$ and $\ker(A)$ to mean the same thing as $\operatorname{image}(T)$ and $\ker(T)$. Many other boooks refer to these sets as the *column space* and *nullspace* of A, respectively. The importance of $\ker(A)$ for solving linear systems is seen in the following fact.

Proposition 6.4. *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$, $\mathbf{x}_p \in \mathbb{R}^n$ be vectors satisfying $A\mathbf{x}_p = \mathbf{b}$. Then a(nother) vector $\mathbf{x} \in \mathbb{R}^n$ solves $A\mathbf{x} = \mathbf{b}$ if and only if*

$$\mathbf{x} = \mathbf{x}_p + \mathbf{x}_h$$

for some vector $\mathbf{x}_h \in \ker(T)$.

The 'p' and 'h' subscripts here stand for 'particular' (as in 'some particular solution of $A\mathbf{x} = \mathbf{b}$') and 'homogeneous' (as in 'solution of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$'.

Definition 6.5. A *linear relation* among vectors $\mathbf{v}_1, \ldots, vv_k \in \mathbb{R}^n$ is a list of scalars $c_1, \ldots, c_k \in \mathbb{R}$ such that

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}.$$

One *always* has the *trivial relation*

$$0\mathbf{v}_1 + \cdots + 0\mathbf{v}_k = \mathbf{0}.$$

The point is that when there is a non-trivial relation among $\mathbf{v}_1, \dots, \mathbf{v}_k$, then one of these vectors can be written as a linear combination of the others. We can remove this vector from the list without changing the span of $\mathbf{v}_1, \dots, \mathbf{v}_k$.

If $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}$, then a linear relation among $\mathbf{v}_1, \dots, \mathbf{v}_k$ is a solution $\mathbf{x} = \begin{pmatrix} c_1 \\ \vdots \\ c_k \end{pmatrix}$ of $A\mathbf{x} = \mathbf{0}$. Hence there are non-trivial relations if and only if $\ker(A)$ contains something other than the zero vector.

Definition 6.6. Vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are called *linearly independent* if

$$c_1\mathbf{v}_1 + \cdots + c_k\mathbf{v}_k = \mathbf{0}$$

only when $c_1 = \cdots = c_k = 0$.

In other words, $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent if there are no non-trivial linear relations among them. If, on the other hand, there exists a non-trivial relation, we say that $\mathbf{v}_1, \dots, \mathbf{v}_k$ are *linearly dependent*.

Remark 6.7. A single vector $\mathbf{v} \in \mathbb{R}^n$ is linearly independent if and only if $\mathbf{v} \neq \mathbf{0}$. Two vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ are linearly independent if and only if neither is a scalar multiple of the other. In general, to check whether $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are linearly independent, you can do the following.

(1) Create the matrix $A = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}$.
(2) Use row operations to find a row equivalent matrix $\tilde{A}$ in row echelon form.
(3) Then $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent if and only if every column of $\tilde{A}$ has a pivot.

In fact, if you discard all the columns of A that correspond to columns of $\tilde{A}$ that lack pivots, then the columns that remain are linearly independent, and they still span image(A).

Remark 6.8. Recall that vectors in the kernel of a matrix A (i.e. the solutions of $A\mathbf{x} = \mathbf{0}$) can be expressed as a linear combination of vectors whose coefficients are the free variables of the RREF matrix $\tilde{A}$ that is row equivalent to A. These vectors are *always* linearly independent.

Definition 6.9. A set of vectors $W \subset \mathbb{R}^n$ is a *subspace* if

(1) $\mathbf{0} \in W$;
(2) if $\mathbf{v} \in W$ and $c \in \mathbb{R}$, then $c\mathbf{v} \in W$;
(3) if $\mathbf{u}, \mathbf{v} \in W$, then $\mathbf{u} + \mathbf{v} \in W$.

Alternatively, a subspace of $\mathbb{R}^n$ is a non-empty set of vectors $W \subset \mathbb{R}^n$ that is closed with respect to scalar multiplication and vector addition.

Example 6.10. Some instances of subspaces of $\mathbb{R}^n$:

- the *trivial* subspace $W = \{\mathbf{0}\}$;
- $W = \mathbb{R}^n$;
- the kernel of a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ (i.e. of an $m \times n$ matrix).
- the image of a linear transformation $T : \mathbb{R}^m \to \mathbb{R}^n$ (i.e. of an $n \times m$ matrix).

Proposition 6.11. *For any* $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$*, the set* $W = \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ *is a subspace of* $\mathbb{R}^n$*.*

Definition 6.12. Let $W \subset \mathbb{R}^n$ be a subspace. A *basis* for W is a linearly independent list of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in W$ that span W.

Example 6.13. The standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n \in \mathbb{R}^n$ constitute a basis for $\mathbb{R}^n$. On the other hand, the trivial subspace $\{\mathbf{0}\} \subset \mathbb{R}^n$ has no basis.

One should note that a basis for a subspace is never unique. If you can find one, you can find many others. The next theorem is probably the most important basic result in linear algebra.

Theorem 6.14 (*Fundamental Theorem of Linear Algebra*)**.** *Every non-trivial subspace of $\mathbb{R}^n$ has a basis, and any two bases for the same subspace have the number of vectors.*

We adopt the convention that the trivial subspace $\{\mathbf{0}\} \subset \mathbb{R}^n$ has dimension 0.
The following auxiliary result is the key ingredient in the proof of Theorem 6.14

Lemma 6.15. *Let $W \subset \mathbb{R}^n$ be a subspace. If $\mathbf{v}_1, \ldots, \mathbf{v}_k \in W$ span W and $\mathbf{u}_1, \ldots, \mathbf{u}_\ell \in W$ are linearly independent, then $k \geq \ell$.*

Corollary 6.16. *Let $W \subset \mathbb{R}^n$ be a subspace. The following are equivalent for a subset $\mathcal{B} = \{\mathbf{v}_1, \ldots, \mathbf{v}_k\} \subset W$.*

- *$\mathcal{B}$ is a basis for W;*
- *$k = \dim W$ and $\mathcal{B}$ is linearly independent subset of W;*
- *$k = \dim W$ and $\mathcal{B}$ spans W.*

The second criterion in this corollary is often rephrased a little informally as: '*$\mathcal{B}$ is a maximal linearly independent subset of W*'. Likewise the third criterion can be rephrased: '*$\mathcal{B}$ is a minimal spanning subset of W*.'

Definition 6.17. The *dimension* $\dim W$ of a subspace $W \subset \mathbb{R}^n$ is

- 0 if $W = \{\mathbf{0}\}$ is trivial;
- the number of vectors in a basis for W, if W is non-trivial.

In particular $\dim \mathbb{R}^n = n$ because the standard basis vectors $\mathbf{e}_1, \ldots, \mathbf{e}_n$ constitute a basis $\mathcal{E}$ for $\mathbb{R}^n$. We can now give a better definition of the rank of a matrix. Henceforth the following supersedes Definition 1.8 above.

Definition 6.18. Let $T : \mathbb{R}^n \to \mathbb{R}^m$ be a linear transformation and A be its matrix. Then

- the *rank* of T (or of A) is the dimension of the image of T;
- the *nullity* of T (or of A) is the dimension of the kernel of T.

To see that this definition doesn't outright contradict our old definition of rank, note the following.

Proposition 6.19. *Let A be an $m \times n$ matrix and $\tilde{A}$ be an RREF matrix row equivalent to A. Then*

- $\operatorname{rank} A$ *is the number of (columns with) pivots in $\tilde{A}$;*
- $\operatorname{nullity} A$ *is the number of columns without pivots in $\tilde{A}$ (i.e. the number of free variables).*

The following is an immediate consequence. Even though the kernel of a linear transformation is a subset of its domain whereas the image of the transformation is a subset of its codomain, the fact is that the dimensions of kernel and range are tightly related.

Theorem 6.20 (*The Rank Theorem*)**.** *If $T : \mathbb{R}^n \to \mathbb{R}^m$ is a linear transformation, then*

$$\operatorname{rank} T + \operatorname{nullity} T = n.$$

7. Coordinates

Proposition 7.1. *Let $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ be a basis for a subspace $W \subset \mathbb{R}^n$. Then for any vector $\mathbf{w} \in W$, there are unique scalars $x_1, \dots, x_k \in \mathbb{R}$ such that*

$$\mathbf{w} = x_1\mathbf{v}_1 + \dots + x_k\mathbf{v}_k.$$

We call the scalars $x_1, \dots, x_k$ in this proposition the *coordinates* of $\mathbf{w}$ relative to the basis $\mathcal{B}$ and let

$$[\mathbf{w}]_{\mathcal{B}} := \begin{pmatrix} x_1 \\ \vdots \\ x_k \end{pmatrix} \in \mathbb{R}^k$$

be the associated *coordinate vector*. Note that the coordinates of $\mathbf{w}$ relative to another basis will be different from the coordinates of $\mathbf{w}$ relative to $\mathcal{B}$. Note also that if we set $S_{\mathcal{B}} := \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}$, then $\mathbf{w}$ and its coordinate vector are related by

$$\mathbf{w} = S_{\mathcal{B}}[\mathbf{w}]_{\mathcal{B}}.$$

Finally, if $\mathcal{E} := \{\mathbf{e}_1, \dots, \mathbf{e}_n\} \subset \mathbb{R}^n$ is the 'standard' basis for $\mathbb{R}^n$, note that $[\mathbf{w}]_{\mathcal{E}} = \mathbf{w}$ for all $\mathbf{w} \in \mathbb{R}^n$.

The following is kind of a mouthful, but it turns out to be extremely useful.

Theorem 7.2. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with matrix $A = [T]$ and $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\} \subset \mathbb{R}^n$ be a basis for $\mathbb{R}^n$. Let $S = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$. Then for any $\mathbf{v} \in \mathbb{R}^n$, the $\mathcal{B}$-coordinates of $\mathbf{v}$ and $T(\mathbf{v})$ are related by*

$$[T(\mathbf{v})]_{\mathcal{B}} = B[\mathbf{v}]_{\mathcal{B}},$$

where the $n \times n$ matrix B satisfies

$$B = S^{-1}AS = \begin{bmatrix} [T(\mathbf{v}_1)]_{\mathcal{B}} & \dots & [T(\mathbf{v}_n)]_{\mathcal{B}} \end{bmatrix}.$$

We call the matrix B in this theorem the *matrix for T relative to the basis $\mathcal{B}$* or, for short, just the *$\mathcal{B}$-matrix* of T. The matrix $[T] = A$ is then the *standard* matrix for T. You might catch me writing $[T]_{\mathcal{B}}$ in place of B since that's a common (though non-Bretscher) sort of notation to use in this context.

8. Orthogonality and Subspaces

A key fact about dot products is the following.

Theorem 8.1 (Cauchy-Schwarz inequality)**.** *For any* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$*, we have* $|\mathbf{v} \cdot \mathbf{w}| \leq \|\mathbf{v}\| \, \|\mathbf{w}\|$*.*

Corollary 8.2 (Triangle inequality)**.** *For any* $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$*, we have* $\|\mathbf{v} + \mathbf{w}\| \leq \|\mathbf{v}\| + \|\mathbf{w}\|$*.*

The Cauchy-Schwarz inequality allows us to make the following definition. In particular, it guarantees that the right side of the equation in the proposition is between -1 and 1.

Definition 8.3. The angle between non-zero vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ is the number $\theta \in [0, \pi]$ satisfying $\cos\theta = \frac{\mathbf{v} \cdot \mathbf{w}}{\|\mathbf{v}\| \|\mathbf{w}\|}$

In particular $\theta = \pi/2$ if $\mathbf{v} \cdot \mathbf{w} = 0$, which is consistent with our earlier definition of orthogonal.

Definition 8.4. A set of vectors $\{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subset \mathbb{R}^n$ is

- *orthogonal* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$;
- *orthonormal* if also $\|\mathbf{v}_i\| = 1$ for $i = 1, \dots, k$.

Note that the standard basis $\{\mathbf{e}_1, \dots, \mathbf{e}_n\}$ for $\mathbb{R}^n$ is an orthonormal set.

Proposition 8.5. *An orthogonal set* $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$ *of non-zero vectors is linearly independent.*

Theorem 8.6. *Every non-trivial subspace* $W \subset \mathbb{R}^n$ *has an orthonormal basis.*

The proof of this theorem amounts to giving a recipe, called the *Gram-Schmidt process* for actually building the orthonormal basis. It works as follows. Since W is non-trivial, we know that it has *some* (not necessarily orthonormal or even orthogonal) basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. One turns this into an orthogonal basis $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ vector-by-vector as follows:

$$\begin{aligned}
\mathbf{w}_1 &= \mathbf{v}_1 \\
\mathbf{w}_2 &= \mathbf{v}_2 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_2) \\
\mathbf{w}_3 &= \mathbf{v}_3 - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_3) - \mathrm{proj}_{\mathbf{w}_2}(\mathbf{v}_3) \\
&\vdots \\
\mathbf{w}_k &= \mathbf{v}_k - \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}_k) - \mathrm{proj}_{\mathbf{w}_2}(\mathbf{v}_k) - \cdots - \mathrm{proj}_{\mathbf{w}_{k-1}}(\mathbf{v}_k).
\end{aligned}$$

Linear independence of $\mathcal{B}$ guarantees that each $\mathbf{w}_j$ is non-zero, and one checks by taking the dot product $\mathbf{w}_j \cdot \mathbf{w}_i$ that $\mathbf{w}_j$ is orthogonal to $\mathbf{w}_i$ for all $i < j$. Hence $\{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ is an orthogonal basis. The *orthonormal* basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ for W is the obtained by normalizing: $\mathbf{u}_j = \frac{\mathbf{w}_j}{\|\mathbf{w}_j\|}$.

Definition 8.7. The *orthogonal complement* of a subspace $W \subset \mathbb{R}^n$ is the set

$$W^\perp := \{\mathbf{v} \in \mathbb{R}^n : \mathbf{v} \cdot \mathbf{w} = 0 \text{ for all } \mathbf{w} \in W\}$$

of all vectors in $\mathbb{R}^n$ orthogonal to W.

The orthogonal complement of the trivial subspace $\{\mathbf{0}\} \subset \mathbb{R}^n$ is all of $\mathbb{R}^n$. Conversely, $(\mathbb{R}^n)^\perp = \{\mathbf{0}\}$. Given a non-zero vector $\mathbf{v} \in \mathbb{R}^2$, the orthogonal complement of the line $\mathrm{span}(\mathbf{v})$ is the line $\mathrm{span}(\mathbf{w})$ where $\mathbf{w}$ is any non-zero vector orthogonal to $\mathbf{v}$.

Proposition 8.8. *Let* $W \subset \mathbb{R}^n$ *be a subspace and* $W^\perp \subset \mathbb{R}^n$ *be its orthogonal complement. Then all of the following are true.*

- *If* $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ *are vectors spanning* W*, then* $\mathbf{w} \in W^\perp$ *if and only if* $\mathbf{w} \cdot \mathbf{w}_1 = \cdots = \mathbf{w} \cdot \mathbf{w}_k = 0$*.*
- $W^\perp$ *is a subspace.*

Theorem 8.9 (*Orthogonal Decomposition Theorem*)**.** *Let* $W \subset \mathbb{R}^n$ *be a subspace. Given any vector* $\mathbf{v} \in \mathbb{R}^n$*, there are unique vectors* $\mathbf{v}_\parallel \in W$ *and* $\mathbf{v}_\perp \in W^\perp$ *such that* $\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp$*. If* $\mathcal{B} = \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ *is an orthogonal basis of* W*, then*

$$\mathbf{v}_\parallel = \mathrm{proj}_{\mathbf{w}_1}(\mathbf{v}) + \cdots + \mathrm{proj}_{\mathbf{w}_k}(\mathbf{v}).$$

We call the vector $\mathrm{proj}_W(\mathbf{v}) := \mathbf{v}_\parallel$ in this theorem the *orthogonal projection of* $\mathbf{v}$ *onto* W.

Remark 8.10. The formula for $\mathbf{v}_\parallel$ in the Theorem is especially nice if we use an *orthonormal* basis $\{\mathbf{u}_1, \dots, \mathbf{u}_k\}$ for W:

$$\mathbf{v}_\parallel = (\mathbf{v} \cdot \mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{v} \cdot \mathbf{u}_k)\mathbf{u}_k$$

Remark 8.11. The discussion following Theorem 8.6 gives a recipe for constructing orthogonal basis vectors $\mathbf{w}_1, \dots, \mathbf{w}_k$ for a subspace W with basis $\{\mathbf{v}_1, \dots, \mathbf{v}_k\}$. We can now rewrite the formula for $\mathbf{w}_j$ more compactly as follows. If $W_{j-1} := \mathrm{span}(\mathbf{v}_1, \dots, \mathbf{v}_{j-1})$, then $\mathbf{w}_j = \mathbf{v}_j - \mathrm{proj}_{W_{j-1}}(\mathbf{v}_j)$.

Theorem 8.9 has a couple of useful (and fairly intuitive) consequences.

Corollary 8.12. *For any subspace* $W \subset \mathbb{R}^n$ *we have*
- $(W^\perp)^\perp = W$*; and*
- $\dim W + \dim W^\perp = n$*.*

9. Transposes and dot products

Definition 9.1. The *transpose* of an $m \times n$ matrix $A = (a_{ij})$ is the $n \times m$ matrix A^T whose ij-entry is a_{ji}.

Another way to say it is that the rows of A^T are the columns of A and vice versa. The reason transposes are important in linear algebra is the following fact about dot products.

Proposition 9.2. *If A is an $m \times n$ matrix, then for any vectors $\mathbf{v} \in \mathbb{R}^m$ and $\mathbf{w} \in \mathbb{R}^n$ we have that*

$$\mathbf{v} \cdot (A\mathbf{w}) = (A^T\mathbf{v}) \cdot \mathbf{w}.$$

Here is a list of the key algebraic properties of transposes.

Proposition 9.3. *If A, B are matrices and $c \in \mathbb{R}$ is a scalar, then*

- $(A^T)^T = A$;
- $(cA)^T = cA^T$;
- $(A + B)^T = A^T + B^T$ *(provided A and B have the same size);*
- $(AB)^T = B^T A^T$ *(provided the number of columns of A equals the number of rows of B).*

Often enough, one encounters square matrices that are equal to their own transposes.

Definition 9.4. A square matrix A is *symmetric* if $A^T = A$.

Note that for *any* matrix $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$, the product $A^T A$ is the $n \times n$ symmetric matrix whose ij-entry is $\mathbf{a}_i \cdot \mathbf{a}_j$.

Definition 9.5. A matrix A is *orthogonal* if $A^T A = I$.

Equivalently, A is orthogonal if its columns $\mathbf{a}_1, \dots, \mathbf{a}_n$ are an orthonormal set.

Remark 9.6. Beware that the order of A^T and A is important in the definition of orthogonality. Given a matrix A, the products $A^T A$ and AA^T are both defined, but they are not generally equal to each other. In fact, the only case when $A^T A$ and AA^T even have the same size is when A is a square matrix. In this case $A^T A = I$ means that A is invertible and $A^{-1} = A^T$. Hence *in this case only* $AA^T = AA^{-1} = I$, too. That is, a *square* matrix A is orthogonal if and only A^T is also orthogonal. You should take a second to appreciate the strangeness of this fact. Another way to state it is that the columns of a square matrix form an orthonormal set if and only if the rows also form an orthonormal set.

The matrix for rotation by θ in $\mathbb{R}^2$ is orthogonal. So is the matrix for reflection about a line through the origin in $\mathbb{R}^2$. Geometrically speaking, the reason one cares about orthogonal matrices is that, like rotations and reflections, they preserve lengths of vectors and angles between vectors.

Theorem 9.7. *If A is an $m \times n$ orthogonal matrix, then for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$, we have $(A\mathbf{v}) \cdot (A\mathbf{w}) = \mathbf{v} \cdot \mathbf{w}$. Hence also*

- $\|A\mathbf{v}\| = \mathbf{v}$;
- *the angle between $A\mathbf{v}$ and $A\mathbf{w}$ is the same as the angle between $\mathbf{v}$ and $\mathbf{w}$.*

The next observation about transposes is important for the discussion in the next section of least square solutions of (inconsistent) linear systems.

Proposition 9.8. *For any matrix A, we have* $(\ker A)^\perp = \operatorname{image} A^T$ *and* $(\operatorname{image} A)^\perp = \ker A^T$*.*

Recall that the image of a matrix is just the span of its columns. Hence $\operatorname{image} A$ is often called the *column space* of A. Likewise, since the columns of A^T are the rows of A, we have that $\operatorname{image} A^T$ is just the span of the rows of A. So $\operatorname{image} A^T$ is often called the *row space* of A.

Corollary 9.9. *For any matrix A we have* $\operatorname{rank} A = \operatorname{rank} A^T$*.*

In other words the subspaces $\operatorname{image} A$ and $\operatorname{image} A^T$ have the same dimension. Keep in mind that if A is $m \times n$, then $\operatorname{image} A \subset \mathbb{R}^m$ while $\operatorname{image} A^T \subset \mathbb{R}^n$. So it's maybe a little surprising that their dimensions should be the same.

10. Least Squares Solutions of Linear Systems

In many practical situations, one arrives at linear systems with many more equations than unknowns. These are typically inconsistent. This section presents a next best alternative to finding an actual solution to such systems.

Definition 10.1. Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$ be a vector. We call $\mathbf{x}^* \in \mathbb{R}^n$ a *least squares solution* of the system $A\mathbf{x} = \mathbf{b}$ if the distance between $A\mathbf{x}^*$ and $\mathbf{b}$ is minimal, i.e. if

$$\|A\mathbf{x}^* - \mathbf{b}\| \leq \|A\mathbf{x} - \mathbf{b}\|$$

for all $\mathbf{x} \in \mathbb{R}^n$.

Theorem 10.2. *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$ be a vector. Then $\mathbf{x}^* \in \mathbb{R}^n$ is a* least squares solution *of $A\mathbf{x} = \mathbf{b}$ if and only if $A^T A\mathbf{x}^* = A^T\mathbf{b}$.*

Definition 10.3. The *normal equation* of a linear system $A\mathbf{x} = \mathbf{b}$ is the associated linear system $A^T A\mathbf{x} = A^T\mathbf{b}$.

The proof of Theorem 10.2 uses the following fact about orthogonal projections.

Proposition 10.4. *Let $W \subset \mathbb{R}^n$ be a subspace and $\mathbf{v} \in \mathbb{R}^n$ be a vector. Then $\operatorname{proj}_W(\mathbf{v})$ is the closest vector in W to $\mathbf{v}$. That is, for any $\mathbf{w} \in W$ we have*

$$\|\mathbf{v} - \mathbf{w}\| \geq \|\mathbf{v} - \operatorname{proj}_W(\mathbf{v})\|,$$

and equality holds only if $\mathbf{w} = \operatorname{proj}_W(\mathbf{v})$.

This justifies the following.

Definition 10.5. The *distance* between a vector $\mathbf{v} \in \mathbb{R}^n$ and a subspace $W \subset \mathbb{R}^n$ is the quantity

$$\operatorname{dist}(\mathbf{v}, W) := \|\mathbf{v} - \operatorname{proj}_W(\mathbf{v})\|.$$

Note that to prove Theorem 10.2 one applies Proposition 10.4 with $W := \operatorname{image} A$ and $\mathbf{v} := \mathbf{b}$.

Theorem 10.2 admits a few refinements, which I summarize here as follows.

Theorem 10.6. *Let A be an $m \times n$ matrix and $\mathbf{b} \in \mathbb{R}^m$ be a vector. Then all of the following hold.*

- *The linear system $A\mathbf{x} = \mathbf{b}$ always has a least squares solution $\mathbf{x}^* \in \mathbb{R}^n$.*
- *If $A\mathbf{x} = \mathbf{b}$ is consistent, then every solution of $A\mathbf{x} = \mathbf{b}$ is a least squares solution and vice versa.*
- *If $\mathbf{x}^* \in \mathbb{R}^n$ is a least squares solution of $A\mathbf{x} = \mathbf{b}$, then $\mathbf{y}^* \in \mathbb{R}^n$ is also a least squares solution of $A\mathbf{x} = \mathbf{b}$ if and only if $\mathbf{x}^* - \mathbf{y}^* \in \ker A$.*
- *Hence $A\mathbf{x} = \mathbf{b}$ has infinitely many least squares solutions if and only if the homogeneous system $A\mathbf{x} = \mathbf{0}$ has infinitely many (ordinary) solutions.*

11. Determinants

To any collection of n vectors $\mathbf{v}_1, \dots, \mathbf{v}_n \in \mathbb{R}^n$, we can associate the set

$$P(\mathbf{v}_1, \dots, \mathbf{v}_n) := \{c_1\mathbf{v}_1 + \dots + c_n\mathbf{v}_n \in \mathbb{R}^n : 0 \le c_1, \dots, c_n \le 1\}.$$

This is called the *parallelotope* with sides $\mathbf{v}_1, \dots, \mathbf{v}_n$. Note in particular, that if $n = 2$, then $P(\mathbf{v}_1, \mathbf{v}_2)$ is just the parallelogram with vertices $\mathbf{0}, \mathbf{v}_1, \mathbf{v}_2, \mathbf{v}_1+\mathbf{v}_2$. And when $\mathbf{v}_1 = \mathbf{e}_1, \dots \mathbf{v}_n = \mathbf{e}_n$ are the standard basis vectors $P(\mathbf{e}_1, \dots, \mathbf{e}_n)$ is just an n-dimensional version of a cube with side lengths 1. In any case, we can think of $\mathbf{v}_1, \dots, \mathbf{v}_n$ as rows in a square matrix A and then introduce the notion of 'determinant of A' in order to compute the n-dimensional volume of $P(\mathbf{v}_1, \dots, \mathbf{v}_n)$.

Definition 11.1. Let $M_{n\times n}$ be the set of all $n \times n$ matrices. We call a function $\det : M_{n\times n} \to \mathbb{R}$ a *determinant* if it has the following properties.

- $\det I = 1$.
- Let $A, B \in M_{n\times n}$ be matrices. If B is obtained by multiplying some row of A by a scalar c, then $\det B = c \det C$;
- If instead B is obtained from A by adding a multiple of one row to a different row, then $\det B = \det A$.

Often, especially during computations, one writes $|A|$ instead of $\det A$.

For the next several results, suppose that $\det : M_{n\times n} \to \mathbb{R}$ is a determinant function. Note that since the scalar c can be negative in the second condition, it is possible that a matrix can have a negative determinant. So in fact, we're actually defining 'volume with a sign' rather than a vanilla never-negative volume. Think of it as the cost of keeping things linear.

Proposition 11.2. *Let $A \in M_{n\times n}$ be a matrix.*

- *If some row of A is equal to $\mathbf{0}$, then* $\det A = 0$.
- *If $B \in M_{n\times n}$ is obtained from A by swapping two different rows of A, then* $\det B = -\det A$

We now understand the effect of all three elementary row operations on the determinant of a matrix. Since every square matrix is row equivalent to either the identity (whose determinant is prescribed to be 1) or to a matrix with a row equal to $\mathbf{0}$, it follows that there is only one possible determinant function. So from now on, we can talk about *the* determinant fucntion, rather than *a* determinant function. Note, though, that one still needs to show that the determinant function actually exists.

Proposition 11.3. *The following are equivalent for $A \in M_{n\times n}$.*

- $\det A \neq 0$*;*
- *A is row equivalent to the identity matrix;*
- *the rows (and columns) of A are linearly independent.*
- *A is invertible.*

The next result is a serious labor saver: when using row operations to find the determinant of a matrix, you don't need to go all the way to reduced echelon form.

Proposition 11.4. *If $A \in M_{n\times n}$ is upper triangular, then $detA = a_{11}a_{22}\dots a_{nn}$ is the product of the diagonal entries of A.*

11.1. **Cofactor expansion.** There is another 'recursive' method for computing determinants. It works well for small matrices, matrices with lots of 0's among their entries, and for matrices whose entries are algebraic expressions rather than specific numbers. But for larger matrices, it is extremely time consuming to execute, even for a computer.

To explain the method we consider first a warm-up case. Note that if $A = (a_{11})$ is 1×1 matrix, then $\det A = a_{11}$. Going one dimension up, we have

Proposition 11.5. *Determinants of* 2×2 *matrices are given by the following formula.*

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

Now on to bigger matrices.

Definition 11.6. The ij-minor of an $n \times n$ matrix A is the $(n-1) \times (n-1)$ matrix A_{ij} obtained by deleting the ith row and jth column of A.

Theorem 11.7. *Let* $A = (a_{ij})$ *be an* $n \times n$ *matrix. Then for any* i *between* 1 *and* n *we have*

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

similarly, for any j *between* 1 *and* n *we have*

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

The quantity $(-1)^{i+j} \det A_{ij}$ is called the *ij-cofactor* of A. We call the first formula in Theorem 11.7 the *cofactor expansion* of $\det A$ about row i and the second formula, the cofactor expansion of A about column j. Both formulas express the determinant of an $n \times n$ matrix as sums of determinants of $(n-1) \times (n-1)$ matrices. Applying the theorem again further gives the latter determinants as sums of determinants of $(n-2) \times (n-2)$ matrices, and so on until you get down to determinants of (a very large number of) 2×2 matrices.

It's somewhat remarkable that one can use either rows or columns for cofactor expansions. In particular, since cofactor expansion about the ith row of A is the same as cofactor expansion about the ith column of A^T, we arrive at the following consequence.

Corollary 11.8. *For any square matrix* A*, we have* $\det A = \det A^T$.

Cofactor expansion leads to many other suprising (but not computationally very useful) formulas in linear algebra. I'll give one of them here (a generalization of Proposition 5.9 above) and you can look in the textbook for others (e.g. Cramer's Rule).

Theorem 11.9. *Let* $A \in M_{n \times n}$ *be a matrix and* $C \in M_{n \times n}$ *be the matrix whose* ij*-entry is* $(-1)^{i+j} \det A_{ij}$. *Then*

$$A^{-1} = \frac{C^T}{\det A}.$$

In case anyone asks, the matrix C in the theorem is sometimes called the *adjugate* or *classical adjoint* of the matrix A. These terms don't get used much in math these days, so I don't expect you to remember them. For myself, I'd just call C the *cofactor* matrix for A.

11.2. **Determinants and linear transformations.** The next result is stated a little imprecisely.

Theorem 11.10. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be a linear transformation with matrix A. Then the n-dimensional volume of any set $\Omega \subset \mathbb{R}^n$ is related to the volume of its image $T(\Omega)$ by*

$$\operatorname{Vol} T(\Omega) = |\det A| \operatorname{Vol}(\Omega).$$

This leads one (philosophically at least) to the following remarkable fact about determinants.

Theorem 11.11. *For any $n \times n$ matrices A and B, we have $\det(AB) = \det(A)\det(B)$.*

It should be stressed that there is *no* corresponding formula for sums of matrices. Anyhow, the fact that determinants distribute over matrix multiplication has some interesting consequences.

Corollary 11.12. *Given $A \in M_{n\times n}$, we have*

- $\det A^n = (\det A)^n$*;*
- *if A is invertible, then $\det(A^{-1}) = 1/\det(A)$.*
- *if A is orthogonal, then $\det A = \pm 1$.*
- *if $B \in M_{n\times n}$ is similar to A, then $\det B = \det A$.*

Concerning the last item in the corollary, recall that a linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ has a matrix $[T]_{\mathcal{B}}$ relative to any given basis for $\mathbb{R}^n$. We therefore define $\det T := \det[T]_{\mathcal{B}}$. One might object that if $\mathcal{C}$ is a different basis for $\mathbb{R}^n$, then typically $[T]_{\mathcal{B}} \neq [T]_{\mathcal{C}}$, so this definition seems to depend on which basis we use. Recall, however, that the matrices $[T]_{\mathcal{B}}$ adn $[T]_{\mathcal{C}}$ are at least similar to each other. Hence the corollary tells us that $\det[T]_{\mathcal{B}} = \det[T]_{\mathcal{C}}$.

12. Diagonalization

The terminology in this section depends on concepts introduced in Section 7. Be prepared to have a look back at that section to make sense of this one.

Recall that a matrix $\Lambda \in M_{n\times n}$ is called *diagonal* if it has the form

$$\Lambda = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$$

Definition 12.1. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ is *diagonalizable* if there exists a basis $\mathcal{B}$ for $\mathbb{R}^n$ such that the matrix $[T]_{\mathcal{B}}$ is diagonal.

We say that the basis $\mathcal{B}$ in this definition *diagonalizes* T.

Remark 12.2. A linear transformation $T : \mathbb{R}^n \to \mathbb{R}^n$ with the same domain and codomain is often called a *linear operator.*

Definition 12.3. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear. A non-zero vector $\mathbf{v} \in \mathbb{R}^n$ is called an *eigenvector* for T with *eigenvalue* $\lambda \in \mathbb{R}$ if $T(\mathbf{v}) = \lambda\mathbf{v}$.

Proposition 12.4. *Let $T : \mathbb{R}^n \to \mathbb{R}^n$ be linear, with (standard) matrix $A := [T] \in M_{n\times n}$. The following are equivalent for a basis $\mathcal{B} = \{\mathbf{v}_1, \dots, \mathbf{v}_n\}$ for $\mathbb{R}^n$.*

- $[T]_{\mathcal{B}} = \Lambda := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix}$ *is diagonal.*
- *Each vector $\mathbf{v}_j \in \mathcal{B}$ is an eigenvector for T with eigenvalue λ_j.*
- $A = S_{\mathcal{B}}\Lambda S^{-1}$ *where* $S_{\mathcal{B}} = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_n \end{bmatrix}$.

One can use this proposition to reformulate the definitions above as statements about matrices.

Definition 12.5. A matrix $A \in M_{n\times n}$ is *diagonalizable* if it is similar to to a diagonal matrix.

Definition 12.6. $\mathbf{v} \in \mathbb{R}^n$ is an *eigenvector* with *eigenvalue* $\lambda \in \mathbb{R}$ for a matrix $A \in M_{n\times n}$ if $A\mathbf{v} = \lambda\mathbf{v}$.

The main point of diagonalization is that it makes it easy to iterate T (equivalently, find powers of A).

Proposition 12.7. *Suppose that $T : \mathbb{R}^n \to \mathbb{R}^n$ given by $T(\mathbf{x}) = A\mathbf{x}$ is diagonalized by a basis $\mathcal{B}$ for $\mathbb{R}^n$. Then the k-fold composition (i.e. kth power) $T^k := T \circ \cdots \circ T$ of T is also diagonalized by cB.*

Specifically,

$$[T]_{\mathcal{B}} = \Lambda := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \quad \text{implies that} \quad [T^k]_{\mathcal{B}} = \Lambda^k := \begin{bmatrix} \lambda_1^k & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ & \vdots & \vdots\ddots & \vdots \\ 0 & 0 & \dots & \lambda_n^k \end{bmatrix}$$

Actually *finding* the eigenvalues and eigenvectors of a square matrix can be tricky. The following proposition allows us to find the eigenvalues and eigenvectors of a linear transformation.

Proposition 12.8. *Let $A \in M_{n\times n}$ be a matrix and $\lambda \in \mathbb{R}$ be a scalar. Then the following are equivalent.*

- *λ is an eigenvalue of A;*
- *$\ker(A - \lambda I)$ is non-trivial;*
- *$\det(A - \lambda I) = 0$.*

In any case a non-zero vector $\mathbf{v} \in \mathbb{R}^n$ is an eigenvector of A with eigenvalue λ if and only if $\mathbf{v} \in \ker(A - \lambda I)$.

This proposition leads to some further observations and definitions.

Definition 12.9. Given $A \in M_{n\times n}$ and $\lambda \in \mathbb{R}$, the λ-*eigenspace* of A is $\ker(A - \lambda I)$

In particular, the λ-eigenspace of A is a subspace of $\mathbb{R}^n$. It is non-trivial if and only if λ is an eigenvalue of A.

Definition 12.10. The *characteristic polynomial* of a matrix $A \in M_{n\times n}$ is the the function $\lambda \mapsto \det(A - \lambda I)$.

Cofactor expansion tells us that the characteristic polynomial of A is a polynomial of degree n. Proposition 12.8 tells us that $\lambda \in \mathbb{R}$ is an eigenvalue of $A \in M_{n\times n}$ if and only if λ is a root of the characteristic polynomial of A.

Unfortunately not all square matrices are diagonalizable. So it is useful to have criteria that we can use to recognize diagonalizability. The next result (especially the second one) leads to a good, if not perfectly general, such criterion.

Proposition 12.11. *Given $A \in M_{n\times n}$, suppose that $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are eigenvectors for distinct eigenvalues $\lambda_1, \dots, \lambda_k$ of A. Then $\mathbf{v}_1, \dots, \mathbf{v}_k$ are linearly independent.*

Corollary 12.12. *If the characteristic polynomial of $A \in M_{n\times n}$ has n distinct real roots, then A is diagonalizable.*

Things get a little complicated when the roots of the characteristic polynomial are *not* distinct.

Definition 12.13. Suppose that $\lambda \in \mathbb{R}$ is an eigenvalue of a matrix $A \in M_{n\times n}$.

- The *algebraic multiplicity* of λ is the multiplicity of λ as a root of the characteristic polynomial of A;
- The *geometric multiplicity* of λ is the dimension of the λ-eigenspace of A.

Typically (i.e. in some probabilistic sense), the geometric and algebraic multiplicities are the same for any given eigenvalue of a matrix A. But there are exceptional cases. What is always true is the following.

Theorem 12.14. *If $\lambda \in \mathbb{R}$ is an eigenvalue of a matrix $A \in M_{n\times n}$, then its geometric multiplicity is no larger than its algebraic multiplicity.*

This leads to the following generalization of Proposition 12.11.

Theorem 12.15. *A matrix $A \in M_{n\times n}$ is diagonalizable if and only if all roots of the characteristic polynomial of A are real, and for each root $\lambda \in \mathbb{R}$ the algebraic multiplicty of λ equals its geometric multiplicity as an eigenvalue of A.*

In closing we note the following fact (related to the comment at the end of Section 11 above) about similar matrices.

Theorem 12.16. *If $A, B \in M_{n\times n}$ are similar, then*

- *A and B have the same characteristic polynomial;*
- *hence A and B have the same eigenvalues, each with the same algebraic multiplicty;*
- *additionally, the geometric multiplicity of any eigenvalue is the same for A and B.*

This is pertinent for the following reason. If $T : \mathbb{R}^n \to \mathbb{R}^n$ is linear, then the matrices $[T]_{\mathcal{B}}$ and $[T]_{\mathcal{C}}$ for T relative to different bases are similar to each other. Hence we can find the characteristic polynomial, the eigenvalues and their algebraic and geometric multiplicities by working with any basis we like for $\mathbb{R}^n$. However, since coordinates of a vector vary with the basis the particular eigenvectors of $[T]_{\mathcal{B}}$ will vary with the basis $\mathcal{B}$.

Appendix A. Equivalent statements about matrices

Let A be an $m \times n$ matrix. By viewing A in different ways, e.g. as the coefficient matrix in a linear system, as a list of columns or rows, or as a linear transformation, one can make many different sounding statements about A that actually amount to the same thing.

The following table gives many of these statements. The first column lists various contexts involving A. Each of the last three columns gives a list of logically equivalent statements. Note that the statements in the first column can hold only when A has more rows than columns (i.e. when $m \geq n$), those in the second can hold only when $n \leq m$, and those in the third can hold only when $m = n$. Note also that the statements in the last column represent the best case scenario when all the statements in the previous two columns hold at the same time[2].

RREF(A)	pivot in each column	pivot in each row	I
Linear system $A\mathbf{x} = \mathbf{b}$	At most one solution	At least one solution	Exactly one solution
Columns of A (i.e. rows of A^T)	linearly independent	span $\mathbb{R}^m$	basis for $\mathbb{R}^m$
Rows of A (columns of A^T)	span $\mathbb{R}^n$	linearly independent	basis for $\mathbb{R}^n$
Linear transformation $T(\mathbf{x}) = A\mathbf{x}$	$\ker(T) = \{\mathbf{0}\}$ i.e. T is injective	$\text{image}(T) = \mathbb{R}^m$ i.e. T is surjective	T invertible i.e. T is bijective
rank A (=rank A^T)	n	m	$n = m$
nullity A	0	$n - m$	$0\ (= n - m)$
nullity A^T	$m - n$	0	$0\ (= m - n)$
det A (only if $m = n$)	N/A	N/A	$\neq 0$

The most important rows in this table are the first four. In any case, you probably shouldn't try to memorize the table so much as understand *why* the various statements in a given column are equivalent to each other. It's mostly a matter of thinking through the definitions of all the words involved, together with a few basic observations (e.g. every non-pivot column of $RREF(A)$ corresponds to a free variable).

[2] I don't think I used the terms (in the linear transformation row) 'injective', 'surjective', 'bijective' in class, so feel free to ignore those if they're unfamiliar.

Appendix B. Proof of the Fundamental Theorem of Linear Algebra

This isn't really needed, but someone asked, so I thought I'd write out the proofs that I gave in class for Lemma 6.15 and Theorem 6.14.

Proof of Lemma 6.15. Since $\mathbf{v}_1, \dots, \mathbf{v}_k$ span W, I can write each of the vectors $\mathbf{w}_1, \dots, \mathbf{w}_\ell \in W$ as linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$. That is, if $V = \begin{bmatrix} \mathbf{v}_1 & \dots & \mathbf{v}_k \end{bmatrix}$, then for each j between 1 and ℓ, I have a vector $\mathbf{a}_j \in \mathbb{R}^k$ such that $V\mathbf{a}_j = \mathbf{w}_j$. In other words

$$VW = A$$

where $W = \begin{bmatrix} \mathbf{w}_1 & \dots & \mathbf{w}_\ell \end{bmatrix}$ and $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_\ell \end{bmatrix}$. Note in particular that A is a $k \times \ell$ matrix.

I claim that $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $A\mathbf{x} = \mathbf{0}$. This implies, among other things, that A is row equivalent to an RREF matrix $\tilde{A}$ with a pivot in each column. Since each row of $\tilde{A}$ has at most one pivot, I infer that $\tilde{A}$ (also a $k \times \ell$ matrix) has at least as many rows as columns. That is, $k \geq \ell$, which is what I wanted to prove.

I still need to justify my claim: if $A\mathbf{x} = \mathbf{0}$, then

$$W\mathbf{x} = VA\mathbf{x} = V\mathbf{0} = \mathbf{0},$$

too. But I can rewrite this as

$$x_1\mathbf{w}_1 + \dots + x_\ell\mathbf{w}_\ell = \mathbf{0}.$$

By hypothesis $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ are linearly independent vectors, admitting only the trivial relation among them. So $x_1 = \dots = x_\ell = 0$; i.e. $\mathbf{x} = \mathbf{0}$, as claimed. □

Now for the main result.

Proof of Theorem 6.14. Since W is non-trivial, there is a non-zero vector $\mathbf{v} \in W$. Hence W contains at least one linearly independent subset $\{\mathbf{v}\}$. On the other hand by Lemma 6.15, *no* linearly independent subset of $\mathbb{R}^n$ has more than n vectors (because $\mathbf{e}_1, \dots, \mathbf{e}_n$ span $\mathbb{R}^n$). This means that we can choose a linearly independent subset $\{\mathbf{v}_1, \dots, \mathbf{v}_k\} \subset W$ with $k \leq n$ as large as possible. I claim that any other vector $\mathbf{w} \in W$ is a linear combination of $\mathbf{v}_1, \dots, \mathbf{v}_k$. That is, $\mathbf{v}_1, \dots, \mathbf{v}_k$ also span W and therefore constitute a basis for W.

In order to justify the claim, let $\mathbf{w} \in W$ be any other vector. Then $\mathbf{v}_1, \dots, \mathbf{v}_k, \mathbf{w}$ is a list of more than k vectors and must therefore be dependent. I.e. there is a non-trivial relation

$$c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k + c_{k+1}\mathbf{w} = \mathbf{0}.$$

If $c_{k+1} = 0$, then we get a simpler relation

$$c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k = \mathbf{0}.$$

But then $c_1 = \dots = c_k = 0$, too, because $\mathbf{v}_1, \dots, \mathbf{v}_k$ are independent. But *one* of the constants $c_1, \dots, c_{k+1}$ must be non-zero, so this can't happen. The alternative is that $c_{k+1} \neq 0$ after all. In this case, I can solve:

$$\mathbf{w} = -\frac{1}{c_{k+1}}(c_1\mathbf{v}_1 + \dots + c_k\mathbf{v}_k).$$

In other words $\mathbf{w} \in \text{span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$, which justifies my claim.

Now that I have one basis for W, suppose $\{\mathbf{w}_1, \dots, \mathbf{w}_\ell\} \in W$ is another. Then Lemma 6.15 tells me that

- $\ell \leq k$ because $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ are independent, whereas $\mathbf{v}_1, \dots, \mathbf{v}_k$ span W; and
- $k \leq \ell$ because $\mathbf{v}_1, \dots, \mathbf{v}_k$ are independent, whereas $\mathbf{w}_1, \dots, \mathbf{w}_\ell$ span W.

So $k = \ell$. That is, any two bases for W consist of the same number of vectors. $\square$

Appendix C. Some things about orthogonality I omitted in class

There are two things concerning orthogonality that I'd hoped to do in class but didn't get to. So I'll put them both here.

Proof of Corollary 8.12. I'll prove $(W^\perp)^\perp = W$ by arguing separately that $W \subset (W^\perp)^\perp$ and then also that $(W^\perp)^\perp \subset W$.

If $\mathbf{w} \in W$, then (by definition of orthogonal complement) $\mathbf{v}\cdot\mathbf{w} = 0$ for all vectors $\mathbf{v} \in W^\perp$. Hence (again by definition) $\mathbf{w} \in (W^\perp)^\perp$. So $W \subset (W^\perp)^\perp$. On the other hand, if $\mathbf{w} \in (W^\perp)^\perp$, then the orthogonal decomposition theorem tells me that there are unique vectors $\mathbf{w}_\parallel \in W$ and $\mathbf{w}_\perp \in W^\perp$ such that $\mathbf{w} = \mathbf{w}_\parallel + \mathbf{w}_\perp$. I claim that $\mathbf{w}_\perp = \mathbf{0}$, so that $\mathbf{w} = \mathbf{w}_\parallel \in W$. Hence $(W^\perp)^\perp \subset W$, too.

To see that the claim is true, observe that $\mathbf{w} \in (W^\perp)^\perp$ and $\mathbf{w}_\perp \in W^\perp$ means that

$$0 = \mathbf{w}_\perp \cdot \mathbf{w} = \mathbf{w}_\perp \cdot (\mathbf{w}_\parallel + \mathbf{w}_\perp) = \mathbf{v}_\perp \cdot \mathbf{w}_\parallel + \|\mathbf{w}_\perp\|^2 = \|\mathbf{w}_\perp\|^2 .$$

Hence $\|\mathbf{w}_\perp\| = 0$, and therefore $\mathbf{w}_\perp = \mathbf{0}$ as claimed.

It remains to prove that $\dim W + \dim W^\perp = n$. Theorem 8.6 tells us that we have an orthogonal basis $\{\mathbf{w}_1, \ldots, \mathbf{w}_k\} \subset \mathbb{R}^n$ for W and an orthogonal basis $\{\mathbf{v}_1 + \ldots \mathbf{v}_\ell\} \subset \mathbb{R}^n$ for $W^\perp$. I claim that the union $\mathcal{B} = \{\mathbf{w}_1, \ldots vw_k, \mathbf{v}_1, \ldots, \mathbf{v}_\ell\}$ is a basis for $\mathbb{R}^n$. Thus $n = k + \ell = \dim W + \dim W^\perp$.

To see that the claim is true, note that (by definition of orthogonal complement) every vector $\mathbf{w}_i$ in the first basis is orthogonal to every vector $\mathbf{v}_j$ in the second basis. Hence $\mathcal{B}$ is an orthogonal set of non-zero vectors and therefore linearly independent. On the other hand, to see that $\operatorname{span}\mathcal{B} = \mathbb{R}^n$, let $\mathbf{v} \in \mathbb{R}^n$ be any given vector. Theorem 8.9 tells me that $\mathbf{v} = \mathbf{v}_\parallel + \mathbf{v}_\perp$, where $\mathbf{v}_\parallel \in W$ and $\mathbf{v}_\perp = W^\perp$. But then $\mathbf{v}_\parallel$ must be a linear combination of $\mathbf{w}_1, \ldots, \mathbf{w}_k$ and $\mathbf{v}_\perp$ must be a linear combination of $\mathbf{v}_1, \ldots, \mathbf{v}_\ell$, so $\mathbf{v}$ itself lies in the span of $\mathcal{B}$. In short, $\mathcal{B}$ is a linearly independent set of vectors that spans $\mathbb{R}^n$, i.e. $\mathcal{B}$ is a basis for $\mathbb{R}^n$ as claimed. □

Next I want to derive a formula for the matrix for orthogonal projection onto a subspace.

Theorem C.1. *Let $W \subset \mathbb{R}^n$ be a basis and $Q = [\mathbf{u}_1, \ldots \mathbf{u}_k]$ be a matrix whose columns form an orthonormal basis for W. Then the matrix for orthogonal projection onto W is given by*

$$[\mathrm{proj}_W] = QQ^T.$$

Proof. If $\mathbf{v} \in \mathbb{R}^n$, then Theorem 8.9 and the fact that $\mathbf{u}_1 \cdot \mathbf{u}_1 = \cdots = \mathbf{u}_k \cdot \mathbf{u}_k = 1$ tell me that

$$\begin{aligned} \mathrm{proj}_W(\mathbf{v}) &= \mathrm{proj}_{\mathbf{u}_1}(\mathbf{v}) + \ldots \mathrm{proj}_{\mathbf{u}_k}(\mathbf{v}) \\ &= (\mathbf{v}\cdot\mathbf{u}_1)\mathbf{u}_1 + \cdots + (\mathbf{v}\cdot\mathbf{u}_k)\mathbf{u}_k \\ &= Q\begin{pmatrix} \mathbf{v}\cdot\mathbf{u}_1 \\ \vdots \\ \mathbf{v}\cdot\mathbf{u}_k \end{pmatrix} = Q\begin{pmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_k^T \end{pmatrix}\mathbf{v} = QQ^T\mathbf{v}. \end{aligned}$$

□

To illustrate this theorem, let $W \subset \mathbb{R}^3$ be the subspace spanned by $\begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix}$ and $\begin{pmatrix} 1 \\ 2 \\ 0 \end{pmatrix}$. Applying the Gram-Schmidt process to these two vectors gives me an orthonormal basis

$\mathbf{u}_1 = \frac{1}{\sqrt{2}}\begin{pmatrix}1\\0\\1\end{pmatrix}$, $\mathbf{u}_2 = \begin{pmatrix}0\\1\\0\end{pmatrix}$. From this we obtain the matrix for orthogonal projection onto W:

$$[\text{proj}_W] = \begin{bmatrix}1/\sqrt{2} & 0\\ 0 & 1\\ 1/\sqrt{2} & 0\end{bmatrix}\begin{bmatrix}1/\sqrt{2} & 0 & 1/\sqrt{2}\\ 0 & 1 & 0\end{bmatrix} = \begin{bmatrix}1/2 & 0 & 1/2\\ 0 & 1 & 0\\ 1/2 & 0 & 1/2\end{bmatrix}.$$

Appendix D. Determinants and matrix multiplication

In class I gave some geometric motivation for the fact that $\det(AB) = (\det A)(\det B)$, but I didn't really prove it. So I'll give a proof here.

Its based on two ideas. First, determinants are unique: i.e. there is only one function det that satisfies all the conditions of Definition 11.1 above. Second, any row operation on a matrix A can be performed by multiplying A on the left by an appropriate matrix E. I need to elaborate this second idea.

Definition D.1. An *elementary matrix* $E \in M_{n\times n}$ is one obtained by performing a single elementary row operation on the identity matrix I_n.

So for instance, the elementary matrix corresponding to multiplying the last row of a 3×3 matrix by 5 is

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

and the elementary matrix corresponding to adding -6 times the 3rd row of a 4×4 matrix to the first is

$$\begin{bmatrix} 1 & 0 & -6 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

Proposition D.2. *Let $A \in M_{m\times n}$ be a matrix and $E \in M_{m\times m}$ be the elementary matrix associated to some row operation. Then the product EA equals the matrix obtained by performing this row operation on A.*

The proof of this amounts to a case-by-case consideration of the elementary matrices corresponding the each of the three types of row operations. I omit the details, noting only that if $A = \begin{bmatrix} \mathbf{a}_1 & \dots & \mathbf{a}_n \end{bmatrix}$, then $EA = \begin{bmatrix} E\mathbf{a}_1 & \dots & E\mathbf{a}_n \end{bmatrix}$. Hence it's enough to check for any given column vector $\mathbf{a}_j$ that $E\mathbf{a}_j$ is the vector obtained from $\mathbf{a}_j$ by performing the row operation on $\mathbf{a}_j$.

Proof of Theorem 11.11. Fix a matrix $B \in M_{n\times n}$. Suppose first that $\det B \neq 0$ and consider the function $D : M_{n\times n} \to \mathbb{R}$ given by $D(A) = \frac{\det(AB)}{\det B}$. We have first of all that

$$D(I) = \frac{\det(IB)}{\det B} = \frac{\det B}{\det B} = 1.$$

Secondly, if E is the elementary matrix corresponding to multiplying the jth row by a scalar $c \in \mathbb{R}$, then $\tilde{A} := EA$ is the matrix obtained from A by mutliplying the jth row of A by c and EAB is the matrix obtained from AB by multiplying the jth row of AB by c. Hence

$$D(\tilde{A}) = D(EA) = \frac{\det((EA)B)}{\det B} = \frac{\det(E(AB))}{\det B} = \frac{c\det(AB)}{\det B} = cD(A).$$

Third, if $\tilde{A}$ is instead obtained from A by adding a multiple of row i to row j and E is now the elementary matrix for *this* row operation, we similary obtain

$$D(\tilde{A}) = D(EA) = \frac{\det((EA)B)}{\det B} = \frac{\det(E(AB))}{\det B} = \frac{\det(AB)}{\det B} = D(A).$$

In short, I've just shown that D satisfies all three conditions in the definition of determinant. Hence

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

Multiplying both sides of this equation by $\det B$ gives the theorem.

It remains to consider the case when $\det B = 0$. For this I recall that the determinant of a matrix vanishes if and only if its kernel is non-trivial (see the table in one of the other bonus sections). So $\det B = 0$ implies that there is a non-zero vector $\mathbf{v} \in \mathbb{R}^n$ such that $B\mathbf{v} = \mathbf{0}$. But then $AB\mathbf{v} = A\mathbf{0} = \mathbf{0}$, too. That is, $\ker(AB)$ is also non-trivial, and therefore $\det(AB) = 0$, too. So in summary

$$\det(AB) = 0 = (\det A) \cdot 0 = (\det A)(\det B)$$

once again. □

Appendix E. Complex eigenvalues and eigenvectors

Given a matrix $A \in M_{n\times n}$, it can happen that some of the roots of the characteristic polynomial $\det(A - \lambda I)$ are not real, but rather complex numbers $\lambda = a + bi$ with $b \neq 0$. In the discussion above, I have allowed only real numbers λ as eigenvalues, but one doesn't have to make that restriction. In the following definition, the letter $\mathbb{C}$ denotes the set of complex numbers and $\mathbb{C}^n$ denotes the set of vectors $\mathbf{v} = \begin{pmatrix} z_1 \\ \vdots \\ z_n \end{pmatrix}$ whose coordinates z_j are allowed to be complex numbers instead of only real numbers.

Definition E.1. $\mathbf{v} \in \mathbb{C}^n$ is an *eigenvector* with *eigenvalue* $\lambda \in \mathbb{C}$ for a matrix $A \in M_{n\times n}$ if $A\mathbf{v} = \lambda\mathbf{v}$.

It is worth noting that if $A \in M_{n\times n}$ is a matrix with real entries (the only sort we've considered in this class), then complex eigenvalues and eigenvectors come in pairs. That is, if $\mathbf{v} = \mathbf{v}_1 + i\mathbf{v}_2 \in \mathbb{C}^n$ is a complex eigenvector for A with complex eigenvalue $\lambda = a + bi \in \mathbb{C}$, then it's *complex conjugate* $\bar{\mathbf{v}} := \mathbf{v}_1 - i\mathbf{v}_2$ is also an eigenvector with eigenvalue $\bar{\lambda} := a - bi$.

Allowing complex eigenvalues and eigenvectors makes it easier to diagonalize matrices. For instance, Theorem E.2 above becomes the following.

Theorem E.2. *A matrix $A \in M_{n\times n}$ is diagonalizable if and only if each root $\lambda \in \mathbb{C}$ of $\det(A - \lambda I)$ has the same algebraic geometric multiplicity as an eigenvalue of A.*

The trade off is that the matrices diagonalizing A have complex entries. In fact, though, there is a useful and interesting alternative to diagonalization when A has complex eigenvalues. I will describe it here only in the case of 2×2 matrices.

Theorem E.3. *Suppose that $T : \mathbb{R}^2 \to \mathbb{R}^2$ is a linear transformation with standard matrix $[T] = A \in M_{2\times 2}$ and that $\lambda = a + bi$, $b \neq 0$, is a complex eigenvalue of T with an eigenvector $\mathbf{v} = \mathbf{v}_1 + i\mathbf{v}_2 \in \mathbb{C}^2$. Then $\mathcal{B} = \{\mathbf{v}_2, \mathbf{v}_1\}$ (note the order here is backward from what you might expect) is a basis for $\mathbb{R}^2$, and the matrix for T relative to $\mathcal{B}$ is*

$$[T]_{\mathcal{B}} = \begin{bmatrix} a & -b \\ b & a \end{bmatrix} = r \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix},$$

where $r = \sqrt{a^2 + b^2}$ and θ are the polar coordinates for $\lambda = a + bi$.

In other words the standard matrix $A := [T]$ for the linear transformation T in this theorem is similar to a scaling-plus-rotation matrix:

$$A = r \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix}^{-1}.$$

The nth power of A is then given by

$$A^n = r^n \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix} \begin{bmatrix} \cos n\theta & -\sin n\theta \\ \sin n\theta & \cos n\theta \end{bmatrix} \begin{bmatrix} \mathbf{v}_2 & \mathbf{v}_1 \end{bmatrix}^{-1}.$$

Appendix F. The Spectral Theorem

Recall that a matrix $S \in M_{m\times n}$ is called orthogonal if $S^T S = I$; equivalently the columns of S constitute an orthonormal set of vectors. So if a square matrix $A \in M_{n\times n}$ has an orthonormal eigenbasis $\mathbf{u}_1, \ldots, \mathbf{u}_n$, we can use the matrix $S = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$ to diagonalize A:

$$A = S\Lambda S^{-1} = S\Lambda S^T,$$

where Λ denotes the diagonal matrix whose ii-entry is the eigenvalue λ_i for $\mathbf{u}_i$. The point here is that S^T is a lot easier to compute than S^{-1}. Moreover, taking the transpose of the entire equation gives

$$A^T = (S\Lambda S^T)^T = (S^T)^T \Lambda^T S^T = S\Lambda S^T = A.$$

So A must be a symmetric matrix! The so-called 'Spectral Theorem' says that the converse is also true.

Theorem F.1 (Spectral Theorem)**.** *A matrix $A \in M_{n\times n}$ has eigenvectors $\mathbf{u}_1, \ldots, \mathbf{u}_n$ that form an orthonormal basis for $\mathbb{R}^n$ if and only if A is symmetric.*

In particular, if A is a symmetric matrix, then all roots of its characteristic polynomial are real, and each has the same geometric and algebraic multiplicity as an eigenvalue of A.

The Spectral Theorem is a little limited by the fact that it only applies to symmetric matrices (although you'd be surprised how often symmetric matrices show up in practice). Recall, however, that if $A \in M_{m\times n}$ is any (not necessarily even square) matrix, then $A^T A \in M_{n\times n}$ is symmetric. One can show in fact that all eigenvalues of $A^T A$ are non-negative. Let $\lambda_1, \ldots, \lambda_n$ denote the *square roots* of the eigenvalues of $A^T A$. These are known as the *singular values of* A.

Theorem F.2 (Singular Value Decompostion)**.** *For any $A \in M_{m\times n}$ there are orthonormal sets $\{\mathbf{u}_1, \ldots, \mathbf{u}_n\} \subset \mathbb{R}^m$ and $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\} \subset \mathbb{R}^n$ such that*

$$A = U\Sigma V^T$$

where $U = \begin{bmatrix}\mathbf{u}_1 & \ldots & \mathbf{u}_n\end{bmatrix}$, $V = \begin{bmatrix}\mathbf{v}_1 & \ldots & \mathbf{v}_n\end{bmatrix}$ are the associated orthogonal matrices, and Σ is the $n \times n$ diagonal matrix whose entries are the singular values of A.

This theorem is the main ingredient in applications of linear algebra such as certain types of data compression and so-called 'principal component analysis'.