MATH 20580

LINEAR ALGEBRA: TERMINOLOGY AND BASIC FACTS

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.

1. Basic definitions and statements

1.1. Vectors in $\mathbb{R}^n$.

Definition 1.1. A *linear combination* of vectors $\mathbf{v}_1, \dots \mathbf{v}_k \in \mathbb{R}^n$ is a vector $\mathbf{v}$ of the form

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

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

Definition 1.2. The *span* of $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ is the set $\mathrm{Span}(\mathbf{v}_1, \dots, \mathbf{v}_k)$ of all linear combinations of $\mathbf{v}_1, \dots, \mathbf{v}_k$.

Definition 1.3. We say that $\mathbf{v}_1, \dots, \mathbf{v}_k \in \mathbb{R}^n$ are *linearly independent* if

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

only when the coefficients all vanish, i.e. when $c_1 = \cdots = c_k = 0$.

Definition 1.4. A *subspace* is a set of vectors $S \subset \mathbb{R}^n$ such that

- $\mathbf{0} \in S$;
- if $\mathbf{v}$ and $\mathbf{w}$ are vectors in S, then $\mathbf{v} + \mathbf{w}$ is also a vector in S.
- if $\mathbf{v}$ is a vector in S and $c \in \mathbb{R}$ is a scalar, then $c\mathbf{v}$ is also a vector in S.

We call the set $\{\mathbf{0}\} \subset \mathbb{R}^n$ containing only the zero vector the *trivial subspace* of $\mathbb{R}^n$.

Definition 1.5. A *basis* for a subspace $S \subset \mathbb{R}^n$ is a collection of vectors $\mathbf{v}_1, \dots, \mathbf{v}_k \in S$ that is linearly independent and that spans S.

Theorem 1.6. *Any non-trivial subspace $S \subset \mathbb{R}^n$ has a basis. Any two bases for S have the same number of vectors.*

Definition 1.7. The *dimension* of a non-trivial subspace $S \subset \mathbb{R}^n$ is the number of vectors in a basis for S.

By common convention, the dimension of the trivial subspace $\{\mathbf{0}\}$ is said to be 0.

Date: February 2, 2025.

1.2. **Matrices and Linear Transformations.**

Definition 1.8. Let A be an $m \times n$ matrix.

- The *column space* of A is the span $\operatorname{Col} A \subset \mathbb{R}^m$ of the columns of A.
- The *row space* of A is the span $\operatorname{Row} A \subset \mathbb{R}^n$ of the rows of A.
- The *null space* of A is the set $\operatorname{Nul} A \subset \mathbb{R}^n$ of solutions $\mathbf{x}$ of the homogeneous linear system $A\mathbf{x} = \mathbf{0}$.

The *rank* of A is the dimension of of $\operatorname{Col} A$, and the *nullity* of A is the dimension of the null space of A.

Theorem 1.9 (Rank Theorem)**.** *For any matrix A, the row and column spaces of A have the same dimension, and they are related to the dimension of the null space of A by*

$$\textit{rank}(A) + \textit{nullity}(A) = \# \textit{ of columns of } A.$$

Replacing A with A^T just turns rows into columns and vice versa. So we have

Corollary 1.10. *A and A^T have the same rank.*

Definition 1.11. A *linear transformation* is a function $T : \mathbb{R}^n \to \mathbb{R}^m$ that 'commutes with scalar multiplication and vector addition'. That is,

- for any vector $\mathbf{v} \in \mathbb{R}^n$ and any scalar $c \in \mathbb{R}$, we have $T(c\mathbf{v}) = cT(\mathbf{v})$;
- for any vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^m$, we have $T(\mathbf{v} + \mathbf{w}) = T(\mathbf{v}) + T(\mathbf{w})$.

Example 1.12. If A is an $m \times n$ matrix, then the function $T_A : \mathbb{R}^n \to \mathbb{R}^m$ given by $T_A(\mathbf{v}) = A\mathbf{v}$ is a linear transformation. We will call T_A the 'matrix transformation associated to A'.

In some sense (at least in the present context), this is the *only* example of a linear transformation.

Theorem 1.13. *Every linear transformation $T : \mathbb{R}^n \to \mathbb{R}^m$ is a matrix transformation. In fact, if $\mathbf{e}_1, \dots, \mathbf{e}_n$ are the standard unit vectors in $\mathbb{R}^n$, then the matrix for T is given column-wise by*

$$A = [T(\mathbf{e}_1) \dots T(\mathbf{e}_n)].$$

2. Many ways of saying the same thing: linear systems

We have several different ways to present a linear system. Consider for instance the following system of two equations in three unknowns:

$$\begin{aligned} 2x_1 + 3x_2 - x_3 &= -1 \\ x_1 - x_2 + 10x_3 &= 0. \end{aligned}$$

We can cut out all the excess symbols and write it as an augmented matrix

$$\begin{bmatrix} 2 & 3 & -1 & -1 \\ 1 & -1 & 10 & 0 \end{bmatrix}.$$

Or we can make it all about linear combinations of vectors

$$x_1 \begin{bmatrix} 2 \\ 1 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ -1 \end{bmatrix} + x_3 \begin{bmatrix} -1 \\ 10 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Or we can state it as a matrix equation

$$\begin{bmatrix} 2 & 3 & -1 \\ 1 & -1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

Or last of all, defining $T : \mathbb{R}^3 \to \mathbb{R}^2$ to be the matrix transformation $T(\mathbf{x}) = \begin{bmatrix} 2 & 3 & -1 \\ 1 & -1 & 10 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$, we can recast the linear system as a functional equation

$$T(\mathbf{x}) = \begin{bmatrix} -1 \\ 0 \end{bmatrix}.$$

These are all valid ways of saying the same thing but with different emphases. For instance, asking whether the linear system has a solution amounts to asking whether the vector on the right side of the vector equation lies in the span of the vectors on the left. And this is the same as asking, in the functional equation, whether the vector on the right is in the range of the linear transformation T. It's a bit of a challenge to keep all this stuff straight at first, but it really helps going forward.

Here's a recap of the above in general terms: let A be an $m \times n$ matrix with jth column $\mathbf{a}_j \in \mathbb{R}^m$ and ij entry $a_{ij} \in \mathbb{R}$; let $\mathbf{x} = (x_1, \dots, x_n) \in \mathbb{R}^n$ and $\mathbf{b} = (b_1, \dots, v_m) \in \mathbb{R}^m$ be vectors; and let $T : \mathbb{R}^n \to \mathbb{R}^m$ be the linear transformation $T(\mathbf{x}) = A\mathbf{x}$. Then the following equations all say exactly the same thing.

m equations in n unknowns.:

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

augmented matrix:

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

vector equation:

$$x_1\mathbf{a_1} + x_2\mathbf{a_2} + \cdots + x_n\mathbf{a_n} = \mathbf{b}.$$

matrix/vector equation:

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

linear transformation equation:

$$T(\mathbf{x}) = \mathbf{b}.$$

2.1. **General solutions of linear systems.** The distributive law for matrix/vector multiplication gives us a nice fact about solutions of linear systems. Namely, a given a linear system $A\mathbf{x} = \mathbf{b}$ might have zero, one or infinitely many solutions, but if $\mathbf{x} = \mathbf{x}_0$ and $\mathbf{x} = \mathbf{x}_1$ are both solutions, then

$$A(\mathbf{x}_0 - \mathbf{x}_1) = A\mathbf{x}_0 - A\mathbf{x}_1 = \mathbf{b} - \mathbf{b} = \mathbf{0}.$$

So any two solutions of the given system differ by a vector in $\operatorname{Nul} A$, i.e. by a solution of the associated homogeneous system $A\mathbf{x} = \mathbf{0}$. Likewise, if $\mathbf{x} = \mathbf{x}_0$ solves $A\mathbf{x} = \mathbf{b}$ and $\mathbf{x} = \mathbf{x}_1$ solves the associated homogeneous system $A\mathbf{x} = \mathbf{0}$, then $\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_1$ also solves $A\mathbf{x} = \mathbf{b}$.

Proposition 2.1. *Suppose that $A\mathbf{x} = \mathbf{b}$ has a (particular) solution $\mathbf{x} = \mathbf{x}_0$. Then another vector $\mathbf{x}$ solves $A\mathbf{x} = \mathbf{b}$ if and only if*

$$\mathbf{x} = \mathbf{x}_0 + \mathbf{x}_h$$

where $\mathbf{x}_h$ is a vector in the nullspace of A. In particular, if $\operatorname{Nul} A$ is trivial, then there is either zero or one solution of $A\mathbf{x} = \mathbf{b}$. And if $\operatorname{Nul} A$ is not trivial, then there are either zero or infinitely many solutions of $A\mathbf{x} = \mathbf{b}$.

More concretely perhaps, when we use Gaussian elimination to solve a linear system, we arrive at a general solution (if any) expressed as the sum of a fixed vector and an arbitrary linear combination of vectors associated to free variables. The fixed vector is the particular soltuion $\mathbf{x} = \mathbf{x}_0$ (what you get if you set all free variables equal to 0), and the remaining linear combination gives the general form for a vector $\mathbf{x}_h$ in the nullspace of A.

3. Many different ways of saying the same thing: the invertible matrix theorem

Thinking about linear systems from various points of view leads to various ways of saying the same thing in different terms about a matrix. Let A be an $m \times n$ matrix. Suppose that A is row equivalent to the matrix A_{red}, which is in reduced echelon form. The following statements are *all* equivalent to saying that A has rank n.

- Every column of A_{red} has a pivot.
- There are no free variables associated to A_{red}.
- If $\mathbf{b} \in \mathbb{R}^n$ is a constant vector, then the linear system $A\mathbf{x} = \mathbf{b}$ has no more than one solution.

- The homogeneous linear system $A\mathbf{x} = \mathbf{0}$ has only the trivial solution $\mathbf{x} = \mathbf{0}$.
- The columns of A are linearly independent.
- The columns of A form a basis for $\operatorname{Col} A$.
- The nullspace of A is trivial, i.e. $\operatorname{Nul} A = \{\mathbf{0}\}$.
- The nullity of A is 0.

The point is not to memorize that these statements are equivalent but rather to understand *why*, in light of things we've done in class, so that you can move back and forth more easily among them.

Here is a different list of statements, all equivalent to saying that A has rank m.

- Every row of A_{red} has a pivot.
- If $\mathbf{b} \in \mathbb{R}^n$ is a constant vector, then the linear system $A\mathbf{x} = \mathbf{b}$ is consistent; i.e. it has at least one solution.
- The columns of A span $\mathbb{R}^n$.
- The rows of A (and of A_{red}) are linearly independent.
- The rows of A (and of A_{red}) form a basis for $\operatorname{Row} A$

Now if A is a square matrix, of size $n \times n$, then one can merge the two lists into an uber-list (once you set $m = n$ in the second list) whose statements are all equivalent to saying that A has rank n. We can even pad the list out further when A is square, tacking on the statements:

- A_{red} is the $n \times n$ identity matrix.
- A is invertible.

Hence the tendency of books to refer to equivalence of all these statements as *The invertible matrix theorem.*

4. Independent subsets, spanning subsets and bases for subspaces

A fundamental fact in linear algebra is that linearly independent sets of vectors in a given subspace are never larger than sets of vectors that span the subspace. To put it more precisely:

Theorem 4.1. *Suppose that $S \subset \mathbb{R}^n$ is a subspace. If $\mathbf{w}_1, \dots, \mathbf{w}_k \in S$ are linearly independent and $\mathbf{v}_1, \dots \mathbf{v}_\ell \in \mathbb{R}^n$ span S. Then $\ell \geq k$.*

Proof. Since $\mathbf{v}_1, \dots, \mathbf{v}_\ell$ span S, we can express any vector $\mathbf{w} \in S$ as a linear combination

$$\mathbf{w} = c_1\mathbf{v}_1 + \dots c_\ell\mathbf{v}_\ell = V\mathbf{a}$$

where $V = [\mathbf{v}_1 \dots \mathbf{v}_\ell]$ is the $n \times \ell$ matrix whose columns are $\mathbf{v}_1, \dots, \mathbf{v}_\ell$ and $\mathbf{a} = (c_1, \dots, c_\ell) \in \mathbb{R}^\ell$ is the vector of coefficients in the combination. Applying this to each of the vectors $\mathbf{w}_1, \dots, \mathbf{w}_k$, we get coefficient vectors $\mathbf{a}_1, \dots, \mathbf{a}_k \in \mathbb{R}^\ell$ such that $V\mathbf{a}_1 = \mathbf{w}_1, \dots, V\mathbf{a}_k = \mathbf{w}_k$. If we let $A = [\mathbf{a}_1 \dots \mathbf{a}_k]$ be the $\ell \times k$ matrix with columns $\mathbf{a}_j$ and $W = [\mathbf{w}_1 \dots \mathbf{w}_k]$ be the $n \times k$ matrix with columns $\mathbf{w}_j$, then we can write this very concisely as a matrix equation.

$$VA = W.$$

Editorial comment: this is an excellent example of why matrix notation is so great.

Because $\mathbf{w}_1, \dots, \mathbf{w}_k$ are linearly independent, we have

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

only if $c_1 = \dots = c_k = 0$. Letting $\mathbf{x} = (c_1, \dots, c_k) \in \mathbb{R}^k$, we can rephrase this by saying that the only solution of the matrix equation

$$W\mathbf{x} = \mathbf{0}.$$

is the trivial solution $\mathbf{x} = \mathbf{0}$. *Editorial comment: see previous editorial comment.*

Putting everything together, we notice that if $\mathbf{x} \in \mathbb{R}^k$ satisfies $A\mathbf{x} = \mathbf{0}$, then

$$W\mathbf{x} = (VA)\mathbf{x} = V(A\mathbf{x}) = V\mathbf{0} = \mathbf{0}.$$

So $\mathbf{x}$ must equal $\mathbf{0}$. That is, the only solution of $A\mathbf{x} = \mathbf{0}$ is $\mathbf{x} = \mathbf{0}$. So when we row reduce A, we'll end up with no free variables, i.e. we'll have a pivot in every column. Since there is at most one pivot in each row, this means A has at least as many rows as columns. But remember: A is an $\ell \times k$ matrix. So $\ell \geq k$. □

This fundamental fact has (at least) two important consequences.

Corollary 4.2. *Any two bases for a subspace $S \subset \mathbb{R}^n$ contain the same number of vectors.*

Proof. Suppose $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_k\}$ and $\mathcal{V} = \{\mathbf{v}_1, \dots \mathbf{v}_\ell\}$ are both bases for S. Then $\mathcal{W} \subset S$ is independent and $\mathcal{V}$ spans S, so $\ell \geq k$. But also $\mathcal{W}$ spans S and $\mathcal{V} \subset S$ is independent. So $k \geq \ell$. Comining the two inequalities, we get $k = \ell$. □

Corollary 4.3. *Any non-trivial subspace $S \subset \mathbb{R}^n$ has a basis.*

Proof. Let $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_k\} \subset S$ be a linearly independent subset such that the number k is as large as possible. Note that since $\mathcal{W} \subset \mathbb{R}^n$ and $\dim \mathbb{R}^n = n$, we must have $k \leq n$. Also note that since S is non-trivial, there's at least one non-zero vector $\mathbf{w} \in S$. Thus $\{\mathbf{w}\} \subset S$ is an independent set, so we must have $k \geq 1$.

I claim now that since k is maximal, $\mathcal{W}$ spans S. To see this, let $\mathbf{v} \in S$ be any given vector. If $\mathbf{v} = \mathbf{w}_j$ for some j, then certainly $\mathbf{v} \in \operatorname{Span} \mathcal{W}$. Otherwise, the set $\mathcal{V} = \{\mathbf{w}_1, \dots, \mathbf{w}_k, \mathbf{v}\}$

has $k+1$ vectors in it and cannot be linearly independent (it's too big!). So there's a linear combination

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

in which not all of the coefficients $c, c_1, \dots, c_k$ vanish. If $c = 0$, then I also have

$$c_1\mathbf{w}_1 + \dots + c_k\mathbf{w}_k = \mathbf{0},$$

in which case $c_1 = \dots = c_k = 0$, because $\mathbf{w}_1, \dots, \mathbf{w}_k$ are independent. So instead, I must have $c \neq 0$, which allows me to solve for $\mathbf{v}$:

$$\mathbf{v} = -\frac{c_1}{c}\mathbf{w}_1 - \dots - \frac{c_k}{c}\mathbf{w}_k.$$

This means that $\mathbf{v} \in \operatorname{Span}\mathcal{W}$. So $\mathcal{W}$ spans S and must be a basis. □

Let us remark that what's implicit in the argument here are some alternative ways to recognize a basis.

Corollary 4.4. *If $S \subset \mathbb{R}^n$ is a non-trivial subspace and $\mathcal{W} = \{\mathbf{w}_1, \dots, \mathbf{w}_k\} \subset S$ is a maximal linearly independent subset, then $\mathcal{W}$ is a basis for S. Alternatively if $\mathcal{V} = \{\mathbf{v}_1, \dots, \mathbf{v}_\ell\} \subset S$ is a minimal spanning set for S, then $\mathcal{V}$ is a basis for S.*