Review: Basics of Linear Algebra

An Algorithmist's Toolkit (CSCI 2952T)

D Ellis Herskowitz
Brown University

The goal of this is to quickly remind you of the basics of linear algebra, some of which we will use in class. I assume that you have already taken at least one linear algebra class and are familiar with standard notation from linear algebra. I won't provide any proofs but if you're looking for more resources and/or proofs I strongly recommend *Linear Algebra Done Right* by Axler.

1 Subspaces, Span, Independence, Bases and Dimension

We begin with some basic notions related to sets of vectors.

Definition 1.1 (Subspaces) *A set of vectors $V \subseteq \mathbb{R}^n$ is called a subspace if for any $u, v \in V$ and $a, b \in \mathbb{R}$ we have*

$$au + bv \in V.$$

Verify For Yourself

If $V \subseteq \mathbb{R}^n$ is a subspace then 0 (the all zeros vector) is in V.

The span of a set of vectors is the result of all possible linear combinations of those vectors; namely, all possible ways of scaling and then adding together vectors in the set.

Definition 1.2 (Span) *Given a set of vectors $V = \{v_1, v_2, \ldots\} \subseteq \mathbb{R}^n$, the span of V is defined as*

$$\textsc{span}(V) := \left\{ \sum_i a_i v_i : a_1, a_2, \ldots \in \mathbb{R} \right\}$$

Verify For Yourself

The span of any set of vectors is a subspace.

Linear independence tells us whether or not one can reconstruct a vector in a set by linearly combining other vectors in that set.

Definition 1.3 (Linear Independence) *A set of vectors $V = \{v_1, v_2, \ldots\} \subseteq \mathbb{R}^n$ is said to be linearly dependent if there exist $a_1, a_2, \ldots \in \mathbb{R}$ such that*

$$\sum_i a_i v_i = 0.$$

Equivalently, V is dependent if there exist $a_1, a_2, \ldots \in \mathbb{R}$ and some $v_i \in V$ such that

$$v_i = \sum_{j \neq i} a_j v_j.$$

If V is not dependent then we say that it is linearly independent (independent for short).

An important class of independent sets of vectors are bases.

Definition 1.4 (Bases) *Given a set of vectors $V \subseteq \mathbb{R}^n$ and $B \subseteq V$, we say that B is a basis of V if*

$$\textsc{span}(B) = V \qquad \text{and} \qquad B \text{ is independent.}$$

An important fact is that every bases has the same size (i.e. number of vectors in it).

Fact 1.5 *Given a set of vectors $V \subseteq \mathbb{R}^n$ and bases B_1 and B_2 of $\mathbb{R}^n$ we have*

$$|B_1| = |B_2|.$$

A notion closely related to bases is that of dimension.

Definition 1.6 (Dimension) *The dimension of a set of vectors $V = \{v_1, v_2, \dots\} \subseteq \mathbb{R}^n$ is the size of the largest set of linearly independent vectors that it contains, namely*

$$\textsc{dim}(V) := \max_{U \subseteq V : U \, independent} |U|.$$

An important fact is that the dimension of a set is always the size of any one of its bases. Using this fact, try and show the following.

Verify For Yourself

$\textsc{dim}(\mathbb{R}^n) = n.$

Observe that the above means that this sense of dimension captures the usual sense for $\mathbb{R}^n$.

Taking the span of a set of vectors leaves its dimension unchanged, as described by the following.

Verify For Yourself

Given any subset $V \subseteq \mathbb{R}^n$, we have $\textsc{dim}(V) = \textsc{dim}(\textsc{span}(V))$.

2 Linear Functions

The most important class of functions studied in linear algebra are so-called linear functions, defined as follows.

Definition 2.1 (Linear Functions) *A function $f : \mathbb{R}^n \to \mathbb{R}^m$ is said to be linear if*

1. *$f(x + y) = f(x) + f(y)$ for any vectors $x, y \in \mathbb{R}^n$*
2. *$f(c \cdot x) = c \cdot f(x)$ for any vector $x \in \mathbb{R}^n$ and $c \in \mathbb{R}$.*

Definition 2.2 (Range) *Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the range of f is*

$$\textsc{range}(f) := \{f(u) : u \in \mathbb{R}^n\}.$$

Notice that the range is a subset of $\mathbb{R}^m$.

Definition 2.3 (Kernel) *Given a function $f : \mathbb{R}^n \to \mathbb{R}^m$, the kernel of f is*

$$\textsc{ker}(f) := \{u \in \mathbb{R}^n : f(u) = 0\}.$$

Notice that the kernel is a subset of $\mathbb{R}^n$

Verify For Yourself

The range and kernel are both subspaces (and therefore both contain 0).

The following summarizes the intuitive idea that the parts of the input space ($\mathbb{R}^n$) that do not get mapped to interesting parts of the output space ($\mathbb{R}^m$) must be mapped to 0 in the output space.

Fact 2.4 (Fundamental Theorem of Linear Maps) *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function then*

$$\textsc{dim}(\mathbb{R}^n) = \textsc{dim}(\textsc{ker}(f)) + \textsc{dim}(\textsc{range}(f)).$$

2.1 Matrices

Matrices occur in linear algebra because they give a natural and unique way of encoding linear functions. Specifically, every linear function corresponds to a (unique matrix) and vice versa.

Fact 2.5 *If $f : \mathbb{R}^n \to \mathbb{R}^m$ is a linear function, then there exists a (unique) $m \times n$ matrix A s.t.*

$$f(x) = Ax$$

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

Fact 2.6 *If A is an $m \times n$ matrix then there exists a (unique) linear function $f : \mathbb{R}^n \to \mathbb{R}^m$ s.t.*

$$Ax = f(x)$$

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

Given the correspondence between matrices and linear maps, we will often abuse notation and use a linear function and its matrix interchangeably. For example, we will talk about $\textsc{ker}(A)$ by which we really mean $\textsc{ker}(f)$ where f is the linear map corresponding to A.

For a given $m \times n$ matrix A, we let $\textsc{rows}(A) \subseteq \mathbb{R}^n$ be all rows of A treated as vectors in $\mathbb{R}^n$ and we let $\textsc{cols}(A) \subseteq \mathbb{R}^m$ be all columns of A, treated as vectors in $\mathbb{R}^m$.

One nice way of thinking about what a matrix A with columns $c_1, c_2, \ldots c_n \in \mathbb{R}^m$ is doing as a function is it takes in a vector $x \in \mathbb{R}^n$ and then linearly combines its columns using x as the coefficients of this combination. That is,

$$Ax = \sum_{c_j \in \textsc{cols}(A)} x_j c_j$$

where in this sum $x_j \in \mathbb{R}$ and $c_j \in \mathbb{R}^m$. Along these lines, check the following for yourself.

Verify For Yourself

$\textsc{span}(\textsc{cols}(\text{A})) = \textsc{range}(\text{A})$.

The following shows that the row and column spans of a matrix have equal dimension.

Fact 2.7 (Row Rank = Col Rank) *Given a matrix A, we have*

$$\textsc{dim}(\textsc{span}(\textsc{rows}(A))) = \textsc{dim}(\textsc{span}(\textsc{cols}(A))).$$

Given the above equality, we can define the rank of a matrix as follows.

Definition 2.8 (Matrix Rank) *We let* $\textsc{rank}(A) := \textsc{dim}(\textsc{span}(\textsc{rows}(A))) = \textsc{dim}(\textsc{span}(\textsc{cols}(A)))$.

Verify For Yourself

If A is an $m \times n$ matrix then $n = \textsc{rank}(A) + \textsc{dim}(\textsc{ker}(A))$.

If A is an $n \times n$ matrix then we say it is full rank iff $\textsc{rank}(A) = n$ (or equivalently, iff $\textsc{dim}(\textsc{ker}(A)) = 0$).

2.2 Inner Products

One important special case of linear functions that we will study are inner products.

Definition 2.9 (Inner Products) *The inner product of vectors* $a, b \in \mathbb{R}^n$ *is defined as*

$$\langle a, b \rangle := \sum_i a_i b_i$$

Notice that if we fix vector $a \in \mathbb{R}^n$ and let $f(x) := \langle a, x \rangle$ then for every $x \in \mathbb{R}^n$ we have $f(x) = Ax$ for the $1 \times n$ matrix A whose single row is a. It follows by the above that the inner product is a linear function. In particular, we have the following.

Fact 2.10 *For every* $x, y \in \mathbb{R}^n$ *we have*

$$\langle a, x + y \rangle = \langle a, x \rangle + \langle a, y \rangle$$

and for all $x \in \mathbb{R}^n$ *and* $c \in \mathbb{R}$ *we have*

$$\langle a, cx \rangle = c \langle a, x \rangle.$$

We will also often use the fact that the inner product is symmetric.

Fact 2.11 *For any* $x, y \in \mathbb{R}^n$ *we have*

$$\langle x, y \rangle = \langle y, x \rangle.$$

3 Systems of Linear Equations

One of the most important applications of linear algebra is solving systems of linear equations.

Definition 3.1 (System of Linear Equations) *A system of linear equations consists of* $a_1, a_2, \ldots, a_m \in \mathbb{R}^n$ *and* $b \in \mathbb{R}^m$ *and asks whether there exists an* $x \in \mathbb{R}^n$ *such that*

$$x_1a_{11} + x_2a_{12} + x_3a_{13} + \ldots + a_{1n} = b_1$$

and

$$x_1a_{21} + x_2a_{22} + x_3a_{23} + \ldots + a_{2n} = b_2$$

and

$$x_1a_{31} + x_2a_{32} + x_3a_{33} + \ldots + a_{3n} = b_3$$

and

$$\ldots$$

and

$$x_1a_{m1} + x_2a_{m2} + x_3a_{m3} + \ldots + a_{nm} = b_m$$

or equivalently whether there exists an $x \in \mathbb{R}^n$ *such that*

$$\langle x, a_i \rangle = b_i \qquad \text{for all } i \in [m]$$

or equivalently whether there exists an $x \in \mathbb{R}^n$ *such that*

$$Ax = b$$

where A *is the matrix whose rows are* $a_1, a_2, \ldots, a_m$*.*

Observe that if $b = 0$ (the all zeros vector) then x is a solution to $Ax = b = 0$ iff it is in $\textsc{ker}(A)$.

Verify For Yourself

The equivalences in Definition 3.1 are actually equivalent.

A useful characterization of the solution set of a system of linear equations is the following.

Fact 3.2 *Let* $K = \{x : Ax = b\}$ *be all solutions to* $Ax = b$*. Then either* $K = \emptyset$ *or for any* $u \in K$ *we have* $K = u + \textsc{ker}(A)$*.*[1]

The above fact along with the Fundamental Theorem of Linear Maps (Fact 2.4) gives us a nice uniqueness criteria for certain systems of linear equations, as follows.

Verify For Yourself

Given $n \times n$ matrix A, if $Ax = b$ has a solution and $\textsc{rank}(A) = n$ then $Ax = b$ has a unique solution.

3.1 Gaussian Elimination

Gaussian elimination is an algorithm which you likely learned in your linear algebra class to solve systems of linear equations.

Fact 3.3 *Given an* $m \times n$ *matrix* A *and* $b \in \mathbb{R}^m$*, one can (by Gaussian elimination) in poly*(n, m) *output an* $x \in \mathbb{R}^n$ *such that* $Ax = b$ *or correctly decide that there is no such* x*.*

[1] Here $u + \textsc{ker}(A) := \{u + v : v \in \textsc{ker}(A)\}$.

If you remember how Gaussian elimination works and you assume that you can add or multiply arbitrarily-large numbers in constant time then the above is quite straightforward to show.[2]

Gaussian elimination is actually a pretty flexible algorithm. For instance, given n vectors $V \subseteq \mathbb{R}^m$, one can use Gaussian elimination to solve a certain set of linear equations to decide if V is independent or not.

Verify For Yourself

Given n vectors $V \subseteq \mathbb{R}^m$, one can in $\text{poly}(n, m)$ time decide if V is independent.

[2] There is a slightly annoying issue here, however, because as you run Gaussian elimination the numbers in your matrix can get larger and larger and so if you don't assume that you can add or multiply arbitrarily-large numbers in constant time it's not entirely clear how to actually implement Gaussian elimination in $\text{poly}(n, m)$ time. Nonetheless, there is a trick one can use (called "Cramer's Rule") to get around this issue.