Lecture 2: Linear Algebra

Rajat Mittal

IIT Kanpur

We will start with the basics of linear algebra that will be needed throughout this course. That means, we will learn about vector spaces, linear independence, matrices and rank in this lecture. The basics of eigenvalue, eigenvectors and spectral decomposition will be covered later. For the rest of this course, I will explain the basics before the application starts as opposed to giving all the basics in the beginning.

This approach is better because you will see the motivation of discussed concepts almost immediately. On the other hand, we might miss out on a few concepts in the initial applications. It is your responsibility to clarify any term or concept which has not been covered.

This material is mostly taken from Gilbert Strang's book "Linear algebra and its applications". For a detailed introduction to these concepts, please refer to Strang's book or any other elementary book on linear algebra.

Note 1. For simplicity of notation, I will use capital letters for matrices and small letters for vectors.

1 System of linear equations

I believe that most of the areas in mathematics owe their existence to a problem. In the case of linear algebra, one of the central problem is to solve equation $Ax = b$. If the matrix A is $m \times n$ then this equation can also be viewed as m linear equations in n variables.

Exercise 1. What are the dimensions of b and x if we assume them to be vectors?

Let us look at a very simple example,

$$\begin{aligned} 2x + y &= 4 \\ x + y &= 3 \end{aligned}$$

You can deduce that the solution is $x = 1, y = 2$. What about,

$$\begin{aligned} 2x + 2y &= 5 \\ x + y &= 3 \end{aligned}$$

You can again prove that these set of equations does not have a solution. So, before solving $Ax = b$, the first question we should ask is,

When does $Ax = b$ have a solution?

We will develop a theory for this question. You might wonder, why develop a theory when you can answer it just by inspection. The reason is, we want to answer this question when we have thousands of variable and equations. The algorithmic answer to these questions is used widely in many industries today.

The theory originates by looking at those two equations in a different manner. Instead of looking at the set of equations as two rows (linear equations), we will view them as column vectors and their combinations.

$$x \begin{bmatrix} 2 \\ 1 \end{bmatrix} + y \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 4 \\ 3 \end{bmatrix}$$

The system of linear equations question can be framed differently now. Does there exist a combination of two vectors, $\begin{bmatrix} 2 \\ 1 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$, which equals $\begin{bmatrix} 4 \\ 3 \end{bmatrix}$?

1.1 Vector spaces

A vector space is a set of elements closed under addition and scalar multiplication (all linear combinations). In other words, V is a vector space iff

$$\forall\ x, y \in V,\ \alpha, \beta \in \mathbb{R} : \quad \alpha x + \beta y \in V.$$

In particular, it implies that for $x, y \in V$, $x + y$ and αx are members of the vector space.

Note 2. We have defined the scalars (α, β) to be from real numbers. But the vector space can be defined over any field by taking the scalars from that field.

There is a more formal definition with axioms about the binary operations and identity element. But the definition above will provide enough intuition for us. The most common examples for a vector space are $\mathbb{R}^n$, $\mathbb{C}^n$, the space of all $m \times n$ matrices and the space of all functions. We will mostly be concerned with finite vector spaces over real number, $\mathbb{R}^n$, in this course.

A subspace is a subset of a vector space which is also a vector space and hence closed under addition and scalar multiplication. A *span* of a set of vectors S is the set of all possible linear combinations of vectors in S. It forms a subspace.

Exercise 2. Give some examples of subspace of $\mathbb{R}^n$. Prove that a span is a subspace.

Note 3. We will be interested in vector space $\mathbb{R}^n$, but the following concepts are valid for general vector spaces.

1.2 Linear independence

To understand the structure of a vector space, we need to understand how can all the elements of a vector space be generated. Using the definition of the vector space, the concept of linear dependence/independence comes out.

Given a set of vectors $v_1, \cdots, v_n \in V$, they are *linearly dependent*, if vector 0 can be expressed as a linear combination of these vectors.

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0,\ \exists i :\ \alpha_i \neq 0.$$

This implies that at least some vector in the set can be represented as the linear combination of other elements. On the other hand, the set is called *linearly independent* iff

$$\alpha_1 v_1 + \cdots + \alpha_n v_n = 0 \Rightarrow \forall i, \alpha_i = 0$$

Intuitively, if we need to find generators of a vector space, a linearly dependent set is redundant. But a linearly independent set might not be able to generate all the elements of a vector space through linear combinations. This motivates the definition of basis, which is, in essence, a maximal linearly independent set of a vector space.

Definition 1. *Basis: A set S is called a basis of a vector space V iff S is linearly independent and any vector in V can be represented as a linear combination of elements in S.*

Since any element in V can be represented as a linear combination of elements of S. This implies that adding any $v \in V \setminus S$ in S will make it linearly dependent (hence a basis is maximal linearly independent set).

One of the basic theorems of linear algebra says that the cardinality of all the basis sets is always the same and it is called the *dimension* of the vector space. Also given a linearly independent set of V, it can be extended to form a complete basis of V (hint: keep adding linearly independent vectors till a basis is obtained).

There is no mention about the uniqueness of the basis. There can be lot of basis sets for a given vector space.

The span of $k < n$ elements of a basis B_1 of V (dimension n) need not be contained in the span of some k' (even $n-1$) elements of B_2. Consider the standard basis $B = \{e_1, \cdots, e_n\}$ and vector $x = (1, 1, \cdots, 1)^T$. Now x or the space spanned by x is not contained in span of any $n-1$ vectors from B.

1.3 Inner product space

All the examples we discussed above are not just vector spaces but inner product spaces. That means they have an associated inner product. Again we won't go into the formal definition. Intuitively, inner product (dot product for $\mathbb{R}^n$) allows us to introduce the concept of angles, lengths and orthogonality between elements of vector space. We will use $x^T y$ to denote the inner product between x and y.

Definition 2. *Orthogonality: Two elements x, y of vector space V are called* orthogonal *iff $x^T y = 0$.*

Definition 3. *Length: The length of a vector $x \in V$ is defined to be $\|x\| = \sqrt{x^T x}$.*

Using orthogonality we can come up with a simpler representation of a vector space. This requires the definition of *orthonormal basis.*

Definition 4. *A basis B of vector space V is orthonormal iff,*

- *For any two elements $x, y \in B$, $x^T y = 0$,*
- *For all elements $x \in B$, $\|x\| = 1$.*

In this orthonormal basis, every vector can be represented as a usual column vector ($n \times 1$ matrix) with respect to this orthonormal basis. It will have co-ordinates corresponding to every basis vector and operation between vectors like summation, scalar multiplication and inner product will make sense as the usual operation on the column vectors.

Given any basis of a vector space, it can be converted into an orthonormal basis. Start with a vector of the basis and normalize it (make it length 1). Take another vector, subtract the components in the direction of already chosen vectors. Normalize the remaining vector and keep repeating this process. This process always results in an orthonormal basis and is known as *Gram-Schmidt Process.*

2 Linear operators

Given two vector spaces, V and W over $\mathbb{R}$, a *linear* operator $M : V \to W$ is defined as an operator satisfying the following properties.

- $M(x + y) = M(x) + M(y)$.
- $M(\alpha x) = \alpha M(x), \ \forall \alpha \in \mathbb{R}$.

These conditions imply that the *zero* of the vector space V is mapped to the *zero* of the vector space W. Also,

$$M(\alpha_1 x_1 + \cdots + \alpha_k x_k) = \alpha_1 M(x_1) + \cdots + \alpha_k M(x_k)$$

Where $x_1, \cdots, x_k$ are elements of V and α_i's are in $\mathbb{R}$. Because of this linearity, it is enough to specify the value of a linear operator on any basis of the vector space V. In other words, a linear operator is uniquely defined by the values it takes on any particular basis of V.

Let us define the addition of two linear operators as $(M + N)(u) = M(u) + N(u)$. Similarly, αM (scalar multiplication) is defined to be the operator $(\alpha M)(u) = \alpha M(u)$. The space of all linear operators from V to W (denoted $L(V, W)$) is a vector space in itself. The space of linear operators from V to V will be denoted by $L(V)$.

Exercise 3. Given the dimension of V and W, what is the dimension of the vector spaces $L(V, W)$?

2.1 Matrices as linear operators

Given two vector spaces $V = \mathbb{R}^n, W = \mathbb{R}^m$ and a matrix M of dimension $m \times n$, the operation $x \in V \to Mx \in W$ is a linear operation. So, a matrix acts as a linear operator on the corresponding vector space.

To ask the converse, can any linear operator be specified by a matrix?

Let f be a linear operator from a vector space V (dimension n) to a vector space W (dimension m). Suppose $\{e_1, e_2, \cdots, e_n\}$ is a basis for the vector space V. Denote the images of this basis under f as $\{w_1 = f(e_1), w_2 = f(e_2) \cdots, w_n = f(e_n)\}$.

Exercise 4. What is the lower-bound/ upper-bound on the dimension of the vector space spanned by $\{w_1, w_2, \cdots, w_n\}$?

Define M_f to be the matrix with columns $w_1, w_2, \cdots, w_n$. Notice that M_f is a matrix of dimension $m \times n$. It is a simple exercise to verify that the action of the matrix M_f on a vector $v \in V$ is just $M_f v$. Here we assume that v is expressed in the chosen basis $\{e_1, e_2, \cdots, e_n\}$.

Exercise 5. Convince yourself that Mv is a linear combination of columns of M.

The easiest way that M_f acts similar to f is: notice that the matrix M_f and the operator f act exactly the same on the basis elements of V. Since both the operations are linear, they are exactly the same operation. This proves that any linear operation can be specified by a matrix.

The previous discussion does not depend upon the chosen basis. We can pick our favorite basis, and the linear operator can similarly be written in the new basis as a matrix (The columns of this matrix are images of the basis elements). In other words, given bases of V and W and a linear operator f, it has a unique matrix representation.

To compute the action of a linear operator, express $v \in V$ in the preferred basis and multiply it with the matrix representation. The output will be in the chosen basis of W. We will use the two terms, linear operator and matrix, interchangeably in future (the bases will be clear from the context).

For a matrix A, A^T denotes the transpose of the matrix.

Let us look at some simple matrices which will be used later.

- Zero matrix: The matrix with all the entries 0. It acts trivially on every element and takes them to the 0 vector.
- Identity matrix: The matrix with 1's on the diagonal and 0 otherwise. It takes $v \in V$ to v itself.
- All 1's matrix (J): All the entries of this matrix are 1.

Exercise 6. What is the action of matrix J?

2.2 Kernel, image and rank

For a linear operator/matrix (from V to W), the *kernel* is defined to be the set of vectors which map to 0.

$$ker(M) = \{x \in V : Mx = 0\}$$

Here 0 is a vector in space W.

Exercise 7. What is the kernel of the matrix J?

The *image* is the set of vectors which can be obtained through the action of the matrix on some element of the vector space V.

$$img(M) = \{x \in W : \exists y \in V,\ x = My\}$$

Exercise 8. Show that $img(M)$ and $ker(M)$ are subspaces.

Exercise 9. What is the image of J?

Notice that $ker(M)$ is a subset of V, but $img(M)$ is a subset of W. The dimension of $img(M)$ is known as the *rank* of M ($rank(M)$). The dimension of $ker(M)$ is known as the nullity of M ($nullity(M)$). For a matrix $M \in L(V, W)$, by the famous rank-nullity theorem,

$$rank(M) + nullity(M) = dim(V).$$

Here $dim(V)$ is the dimension of the vector space V.

Proof. Suppose $u_1, \cdots, u_k$ is the basis for $ker(M)$. We can extend it to the basis of V, $u_1, \cdots, u_k, v_{k+1}, \cdots, v_n$. We need to prove that the dimension of $img(M)$ is $n - k$. It can be proved by showing that the set $\{Mv_{k+1}, \cdots, Mv_n\}$ forms a basis of $img(M)$.

Exercise 10. Prove that any vector in the image of M can be expressed as linear combination of $Mv_{k+1}, \cdots, Mv_n$. Also any linear combination of $Mv_{k+1}, \cdots, Mv_n$ can't be zero vector.

□

Given a vector v and a matrix M, it is easy to see that the vector Mv is a linear combination of columns of M. To be more precise, $Mv = \sum_i M_i v_i$ where M_i is the ith column of M and v_i is the ith co-ordinate of v. This implies that any element in the image of M is a linear combination of its columns.

Exercise 11. Prove the rank of a matrix is equal to the dimension of the vector space spanned by its columns (column-space).

The dimension of the column space is sometimes referred to as the *column-rank*. We can similarly define the *row-rank*, the dimension of the space spanned by the rows of the matrix. Luckily, row-rank turns out to be equal to column-rank and we will call both of them as the rank of the matrix. This can be proved easily using *Gaussian elimination*. We will give a *visual* proof of the theorem.

Proof. Given an $m \times n$ matrix M, say $\{c_1, c_2, \cdots, c_k\}$ span the column space of M. Suppose, C be the $m \times k$ matrix with columns $\{c_1, c_2, \cdots, c_k\}$. Then, there exist an $k \times n$ matrix R, s.t., $CR = M$. If $\{d_1, d_2, \cdots, d_k\}$ are the columns of R, then the equation $CR = M$ can be viewed as,

$$\begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ c_1 & c_2 & \cdot & \cdot & c_k \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ d_1 & d_2 & \cdot & \cdot & d_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix} = \begin{pmatrix} \vdots & \vdots & \vdots & \vdots & \vdots \\ Cd_1 & Cd_2 & \cdot & \cdot & Cd_n \\ \vdots & \vdots & \vdots & \vdots & \vdots \end{pmatrix}$$

Another way to view the same equation is,

$$C \begin{pmatrix} \cdots & r_1 & \cdots \\ \cdots & r_2 & \cdots \\ \cdots & \cdot & \cdots \\ \cdots & \cdot & \cdots \\ \cdots & r_k & \cdots \end{pmatrix} = \begin{pmatrix} \cdots & \sum_i C_{1i} r_i & \cdots \\ \cdots & \sum_i C_{2i} r_i & \cdots \\ \cdots & \cdot & \cdots \\ \cdots & \cdot & \cdots \\ \cdots & \sum_i C_{ki} r_i & \cdots \end{pmatrix}$$

This shows that the k columns of R span the row-space of M. Hence, column-rank is smaller than the row-rank.

Exercise 12. Show that row-rank is less than column-rank by a similar argument.

□

Note 4. The column-rank is equal to row-rank. It does not mean that the row-space is same as the column-space.

Using these characterizations of rank, it can be proved easily that $rank(A) = rank(A^T)$ and $rank(A + B) \leq rank(A) + rank(B)$.

2.3 $\boldsymbol{Ax = b}$

To conclude, we will again look at the original question about the existence of solutions for a system of linear equations. Given $Ax = b$, there are two cases, depending upon the rank of A.

- Non-singular: If A is full rank (the columns are linearly independent) then there is a unique solution.
- Singular: If A's rank is not full (the columns are linearly dependent) then there could be no solution or infinite solutions.

To test whether the columns are linearly dependent or not (and solve the linear equations), the established method in practice is Gaussian elimination. I assume that you know what Gaussian elimination is, if not, please read about it.

Given an $n \times n$ matrix, Gaussian elimination works in time $O(n^3)$. There are better algorithms known, but Gaussian elimination will be enough for our purposes.

3 Assignment

Exercise 13. When A is singular, under what conditions will we have no solution for the system $Ax = b$?

Exercise 14. Prove that $rank(AB) \leq rank(A)$.

Exercise 15. Read about Gaussian elimination.

Exercise 16. Prove that $rank(A) = rank(A^*A)$.

Hint: $rank(A) \geq rank(A^*A)$ is easy. For the other direction, reduce A to its reduced row echelon form.

Exercise 17. Show that $v^T Aw = \sum_{ij} A_{ij} v_i w_j$, where A is a matrix and v, w are vectors.

Exercise 18. Prove that $Trace(AB) = Trace(BA)$.

Exercise 19. Show that $trace(A(v^T v)) = v^T Av$, where A is a matrix and v is a vector.

References

1. G. Strang. Linear Algebra and Its Applications. *Cengage learning*, 2007.
2. S. Boyd and L. Vandenberghe. Convex Optimization. *Cambridge*, 2004.
3. D. Spielman. Course notes: Spectral Graph Theory. *http://www.cs.yale.edu/homes/spielman/561/*, 2015.
4. L. Trevisan. Course notes: Graph Partitioning, Expanders and Spectral Methods. *https://people.eecs.berkeley.edu/ luca/expanders2016/index.html.*