Lecture 9: Matrices

Rajat Mittal *

IIT Kanpur

We have studied vector space $\mathbb{R}^n$ and various structures in this space till now. A matrix is an object in $\mathbb{R}^{n \times n}$. It can be viewed as a vector in space $\mathbb{R}^m$, where $m = n^2$, with additional structure (rows and columns). The next few lectures will talk about how this extra structure gives rise to various properties.

1 Linear operators

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

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

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

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

Where $x_1, \cdots, x_k$ are elements of V and α_i's are in $\mathbb{R}$. Because of this linearity, if the function is defined on any basis, it can be extended to the entire vector space. Hence a function is uniquely specified by the values it takes on any chosen basis of V.

1.1 Matrices as linear operators

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

Can any linear operation be specified by a matrix. The standard basis for a vector space is $\mathbb{R}^n$ is $e_1, \cdots, e_n$. Where e_i is the vector whose i^{th} co-ordinate is 1 and rest 0.

Suppose $w_1, \cdots, w_n$ are the images of $e_1, \cdots, e_n$ under the linear operation f respectively. Then define M_f to be the matrix with columns $w_1, \cdots, w_n$. Notice that the matrix M_f and f act the same on standard basis. From the observation in the previous section, they are the same operation. Hence any linear operation can be specified as a matrix.

The previous discussion does not depend upon the chosen basis. We can pick our favorite orthonormal basis, and the linear operator can similarly be written in the new basis as a matrix (The columns of matrix are images of orthonormal basis). To compute the action of a linear operator, express $v \in V$ in the preferred basis and multiply with the matrix. We will use these terms interchangeably in future (the basis will be clear from the context).

1.2 Examples of matrices

- Zero matrix: The matrix with all the entries O. 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 are 1.

Exercise 1. What is the action of matrix J.

* Thanks to John Watrous's course notes, IQC, Waterloo.

1.3 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. The *image* is the set of vectors which can be obtained through the action of the matrix.

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

Exercise 2. Show that $img(M$ and $ker(M)$ are subspaces.

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 $M \in L(V)$, 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 $n - k$. For this, it can be shown that $Mv_{k+1}, \cdots, Mv_n$ forms the basis of $img(M)$.

Exercise 3. Prove that any vector in image 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.

□

Exercise 4. What is the kernel and image of J (the all 1's matrix)?

1.4 Vector space of linear operators

The addition of two matrices is defined as the entrywise addition. Similarly αM (scalar multiplication) is defined to be the matrix having all entries of M multiplied by α. Hence the space of all linear operators from V to W (denoted $L(V, W)$) is a vector space. The space of linear operators from V to V will be denoted as $L(V)$.

Exercise 5. What is the dimension of these vector spaces?

In convex optimization, some times the cone we are concerned with will lie in these spaces of linear operators.

2 Operations on matrices

Lets look at some of the basic operations on these matrices.

2.1 Transpose

Given an $m \times n$ matrix A, the *transpose* of this matrix A^T is an $n \times m$ matrix, s.t.,

$$A[i, j] = A^T[j, i].$$

Here $A[i, j]$ denotes the entry in the i^{th} row and j^{th} column. A matrix is called *symmetric* iff it is equal to its transpose,

$$A = A^T.$$

2.2 Trace

The *trace* of a matrix is the sum of all the diagonal elements.

$$tr(A) = \sum_i A[i,i]$$

2.3 Multiplication

Given two matrices A and B, their multiplication AB is defined as,

$$AB[i,j] = \sum_k A[i,k]B[k,j].$$

From the definition it is clear that the multiplication AB makes sense iff the number of columns of A are the same as number of rows of B. This definition is preferred over entrywise multiplication, because it agrees with the notion of composition of linear operators.

2.4 Entrywise multiplication

The entrywise multiplication of two matrices is known as *Hadamard product* and only makes sense when both of them have same number of rows and columns. The Hadamard product of two matrices A, B is

$$(A \circ B)[i,j] = A[i,j]B[i,j].$$

The related operation is when you add up the entries of this Hadamard product.

$$(A \bullet B) = \sum_{i,j} A[i,j]B[i,j]$$

Notice that $A \bullet B$ is a scalar and not a matrix.

Exercise 6. Given a matrix, express $\bullet$ operation in terms of multiplication and trace operation.

2.5 Direct sum

Given matrices $A \in L(V)$ and $B \in L(W)$, their direct sum is in $L(V \oplus W)$.

$$A \oplus B = \begin{pmatrix} A & 0 \\ 0 & B \end{pmatrix}$$

Notice that the dimension of space $V \oplus W$ is $dim(V) + dim(W)$.

2.6 Tensor product

Given matrices $A \in L(V)$ and $B \in L(W)$, their tensor product is in space $L(V \otimes W)$.

$$(A \otimes B)[(i,k),(j,l)] = A[i,j]B[k,l]$$

Exercise 7. What is the dimension of space $V \otimes W$?

2.7 Inverse

Inverse of a matrix M is the matrix M^{-1}, s.t., $MM^{-1} = M^{-1}M = I$. The inverse only exists if the matrix has full rank (the columns of M span the whole space).

Exercise 8. What is the inverse of matrix J (all 1's matrix).

3 Eigenvalues and eigenvectors

A matrix $M \in L(V, w)$ is square if $dim(V) = dim(W)$. In particular, a matrix $M \in L(V)$ is always square. Consider a matrix $M \in L(V)$, any vector $v \in V$ which satisfies $Mv = \lambda v$ for some $\lambda \in \mathbb{R}$ is called the eigenvector of matrix M. λ is called the eigenvalue corresponding to eigenvector v.

Exercise 9. Given two eigenvectors v, w, when is there linear combination an eigenvector itself?

From the previous exercise, all the eigenvectors corresponding to a particular eigenvalue form a subspace. It is called the *eigenspace* of the corresponding eigenvalue.

Any eigenvalue λ of matrix $M(n \times n)$ satisfies the equation $Det(\lambda I - M) = 0$. This is a polynomial of degree n and will have n roots. But these roots might not be real. So the matrix might or might not have n eigenvalues. The situation is much simpler in complex ($\mathbb{C}$) field. Where all the n roots exist and precisely correspond to the n eigenvalues.

The polynomial $Det(\lambda I - M) = 0$ is called the *characteristic polynomial* of M.

Theorem 1. *Given a matrix P of full rank, matrix M and matrix $P^{-1}MP$ have same eigenvalues.*

Proof. Suppose λ is an eigenvalue of $P^{-1}MP$, we need to show that it is an eigenvalue for M too. Say λ is an eigenvalue with eigenvector v. Then,

$$P^{-1}MPv = \lambda v \Rightarrow M(Pv) = \lambda Pv.$$

Hence Pv is an eigenvector with eigenvalue λ.

The opposite direction follows similarly. Given an eigenvector v of M, it can be shown that $P^{-1}v$ is an eigenvector of $P^{-1}MP$.

$$P^{-1}MP(P^{-1}v) = P^{-1}Mv = \lambda P^{-1}v$$

Hence proved. □

Exercise 10. Show that matrix M and M^T have same eigenvalues.