Linear Algebra and Vector Calculus Notes and Exercises

Somnath Sikdar

February 2, 2020

Contents

1	Linear Algebra Basics		2
	1.1	Linear Functions	2
	1.2	Image and Kernel of a Linear Function	2
	1.3	Quadratic Forms	3
		1.3.1 Definiteness	4

Chapter 1

Linear Algebra Basics

1.1 Linear Functions

A function $L: \mathbb{R}^m \to \mathbb{R}^n$ is a linear function if for all $x, y \in \mathbb{R}^m$ and for all $a, b \in \mathbb{R}$

$$L(ax + by) = aL(x) + bL(y).$$

It follows (by induction) that for all $x_1, \ldots, x_r \in \mathbb{R}^m$ and all $a_1, \ldots, a_r \in \mathbb{R}$

$$L(a_1x_1+\cdots+a_rx_r)=a_1L(x_1)+\cdots+a_rL(x_r).$$

Theorem 1.1. A linear function $L: \mathbb{R}^m \to \mathbb{R}^n$ is completely determined by its effect on the standard basis vectors e_1, \ldots, e_m of \mathbb{R}^m . An arbitrary choice of vectors $L(e_1), \ldots, L(e_m)$ of \mathbb{R}^n determines a linear function from \mathbb{R}^m to \mathbb{R}^n .

Proof. Given any vector $x \in \mathbb{R}^m$, we can express it as a unique linear combination $\sum_{i=1}^m \alpha_i e_i$ of the basis vectors. By the linearity of L, $L(x) = \sum_i \alpha_i L(e_i)$ which is completely specified by $L(e_1), \ldots, L(e_m)$.

Let b_1, \ldots, b_m be any vectors in \mathbb{R}^n . Define a map L from \mathbb{R}^m to \mathbb{R}^n as follows: for $x = \sum_{i=1}^m \alpha_i e_i \in \mathbb{R}^m$, $L(x) = \sum_{i=1}^m \alpha_i b_i$. Then $L(e_i) = b_i$ for all $1 \le i \le m$ and for all $x, y \in \mathbb{R}^m$ and all $a, b \in \mathbb{R}$:

$$L(ax + by) = \sum_{i=1}^{m} (ax_i + by_i)b_i = a\sum_i x_i b_i + b\sum_i y_i b_i = aL(x) + bL(y)$$

Note that the domain of definition of a linear function must be a vector space. An non-linear function can be defined on a subset of a vector space.

1.2 Image and Kernel of a Linear Function

The image $\operatorname{Im}(L)$ of a linear function $L: \mathbb{R}^m \to \mathbb{R}^n$ is the set of vectors in \mathbb{R}^n that L maps \mathbb{R}^m to. In symbols, $\operatorname{Im}(L) := \{L(x) \in \mathbb{R}^n : x \in \mathbb{R}^m\}$. The kernel $\operatorname{Ker}(L)$ of L is the set of vectors in \mathbb{R}^m that L maps to the zero vector in \mathbb{R}^n : $\operatorname{Ker}(L) := \{x \in \mathbb{R}^m : L(x) = \mathbf{0}_n\}$.

Theorem 1.2. Let $L: \mathbb{R}^m \to \mathbb{R}^n$ be a linear function. Then the following hold:

- 1. Im(L) is a subspace of \mathbb{R}^n ;
- 2. Ker(L) is a subspace of \mathbb{R}^m ;
- 3. $\dim(\operatorname{Im}(L)) + \dim(\operatorname{Ker}(L)) = \dim(\mathbb{R}^m) = m$.

Proof. The proof of (1) and (2) are similar. For proving (1), let $z_1, z_2 \in \text{Im}(L)$ and $\alpha, \beta \in \mathbb{R}$. Then there exist $x_1, x_2 \in \mathbb{R}^m$ such that

$$L(\alpha x_1 + \beta x_2) = \alpha z_1 + \beta z_2,$$

implying that $\alpha z_1 + \beta z_2 \in \text{Im}(L)$. Since Im(L) is closed under vector addition and multiplication by scalars, it is a vector space. Since it is a subset of \mathbb{R}^n , it must be a subspace of \mathbb{R}^n .

To prove (3), let us assume that $\dim(\operatorname{Ker}(L)) = k$ and that b_1, \ldots, b_k is a basis of $\operatorname{Ker}(L)$. Extend this basis to a basis $b_1, \ldots, b_k, b_{k+1}, \ldots, b_m$ for \mathbb{R}^m . Then for every vector $x \in \mathbb{R}^m$ there exist scalars β_1, \ldots, β_m such that $x = \sum_{i=1}^k \beta_i b_i$. Moreover since L is linear,

$$L(x) = \sum_{i=1}^{m} \beta_i L(b_i)$$

= $\sum_{i=k+1}^{m} \beta_i L(b_i)$.

This shows that every vector in Im(L) can be expressed as a linear combination of the vectors $L(b_{k+1}), \ldots, L(b_m)$. To show that they form a basis of Im(L), it is sufficient to show that they are linearly independent. Suppose not. Then there exist scalars $\beta_{k+1}, \ldots, \beta_m$, not all zero, such that $\sum_{i=k+1}^m \beta_i L(b_i) = \mathbf{0}_n$. By the linearity of L, we have $L(\sum_{i=k+1}^m \beta_i b_i) = \mathbf{0}_n$ and hence $\mathbf{0}_m \neq \sum_{i=k+1}^m \beta_i b_i \in Ker(L)$. This is a contradiction since the vectors b_{k+1}, \ldots, b_m are not in the space spanned by the vectors b_1, \ldots, b_k . Thus the vectors $\beta_{k+1}, \ldots, \beta_m$ must be independent and form a basis of Im(L). This proves (3).

1.3 Quadratic Forms

A function $q: \mathbb{R}^m \to \mathbb{R}$ is called a *quadratic form* if there exists a real symmetric $m \times m$ matrix A such that for all $x \in \mathbb{R}^m$ such that

$$q(x) = x^{\mathsf{T}} A x$$
.

The right-hand side of the equation may be written as:

$$(x_{1},...,x_{m})\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{m1} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \begin{pmatrix} x_{1} \\ \vdots \\ x_{m} \end{pmatrix} = (x_{1},...,x_{m})[x_{1}A_{*1} + \cdots + x_{m}A_{*m}]$$

$$= \sum_{i=1}^{m} \sum_{j=1}^{m} x_{i}a_{ij}x_{j}.$$

The symbol A_{*j} denotes the jth column of A. The name "quadratic" form arises from the last expression which is what a quadratic expression in m variables looks like.

Example 1.1. The function

$$q(h_1, h_2, h_3) = h_1^2 + h_2^2 + h_3^2 + 2h_1h_2 + 2h_2h_3 + 2h_1h_3 = (h_1 + h_2 + h_3)^2$$

can be expressed as $h^{T}Ah$, where A is the all-ones 3 × 3 matrix.

1.3.1 Definiteness

A quadratic form $q: \mathbb{R}^m \to \mathbb{R}$ is

- 1. positive definite if for all non-zero $x \in \mathbb{R}^m$, q(x) > 0;
- 2. *negative definite* if for all non-zero $x \in \mathbb{R}^m$, q(x) > 0;
- 3. indefinite if q(x) takes on both positive and negative values.

Example 1.2. Let $q(h) = -h_1^2 + 2h_2^2 - h_3^2$ such that the matrix associated with q is

$$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

The q(0,1,0) = 2 and q(1,0,1) = -2. Thus the quardatic form q is indefinite.

Example 1.3. The quadratic form $q(h_1, h_2, h_3) = h_1^2 + h_2^2 + 4h_3^2 + 2h_1h_2 - 4h_2h_3 - 4h_1h_3$ is positive definite, since we may re-write $q(h_1, h_2, h_3)$ as $(h_1 + h_2 - 2h_3)^2$ which is strictly positive for all $(h_1, h_2, h_3) \in \mathbb{R}^3 \setminus \{0\}$.

To characterize the definiteness of quadratic forms, we make use of the following properties of real symmetric matrices.

Lemma 1.1. The eigenvectors of a real symmetric matrix A are real.

Proof. Suppose that *A* is an $m \times m$ matrix and suppose that some eigenvalue λ is complex. Let x be a eigenvector corresponding to it, which can be a complex vector. Then

$$Ax = \lambda x. \tag{1.1}$$

Taking complex conjugates on each side, we get:

$$A\bar{x} = \bar{\lambda}\bar{x}.\tag{1.2}$$

Pre-multiply (1.1) by \bar{x}^{\dagger} and (1.2) by x^{\dagger} to obtain:

$$\bar{x}^{\mathsf{T}} A x = \bar{x}^{\mathsf{T}} \lambda x \tag{1.3}$$

$$x^{\mathsf{T}} A \bar{x} = x^{\mathsf{T}} \bar{\lambda} \bar{x}. \tag{1.4}$$

Take the transpose of equation (1.4), we obtain: $\bar{x}^{\dagger}Ax = \bar{x}^{\dagger}\bar{\lambda}x$, where we made use of the fact that *A* is symmetric. Now subtracting this equation from (1.3), we obtain:

$$x^{\mathsf{T}}(\lambda - \bar{\lambda})\bar{x} = 0.$$

Since $x^{\mathsf{T}}\bar{x}$ is the sum of products of complex conjugates, it is not zero unless each component of x is zero. Since this is not the case (x is an eigenvector), we must have $\lambda = \bar{\lambda}$ and hence λ is real.

Example 1.4. Consider the unit matrix

$$I_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}.$$

This has as eigenvalues 1 and 1, which are clearly not distinct. In general, I_m has m eigenvalues all of which are 1. Note that I_m has rank m. On the other hand, the matrix

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$$

has rank one and has two distinct eigenvalues 0,2 which are the roots of the equation: $(1-\lambda)^2-1=0$. Thus the rank of a matrix has nothing to do with the number of distinct eigenvalues.

Now that we know that a real symmetric matrix has only real eigenvalues, what can we say about the corresponding eigenvectors?

Lemma 1.2. The eigenvectors of a real symmetric matrix corresponding to distinct eigenvalues are mutually orthogonal.

Proof. Let λ_i and λ_j be two distinct eigenvalues of a real symmetric matrix A. Let x_i and x_j be eigenvectors corresponding to these eigenvalues. Then $Ax_i = \lambda_i x_i$ and $Ax_j = \lambda_j x_j$. Pre-multiplying the first of these equations by x_i^{T} and the second by x_i^{T} , we obtain:

$$\boldsymbol{x}_{i}^{\mathsf{T}} A \boldsymbol{x}_{i} = \boldsymbol{x}_{i}^{\mathsf{T}} \lambda_{i} \boldsymbol{x}_{i} \tag{1.5}$$

$$\boldsymbol{x}_{i}^{\mathsf{T}} A \boldsymbol{x}_{i} = \boldsymbol{x}_{i}^{\mathsf{T}} \lambda_{i} \boldsymbol{x}_{i} \tag{1.6}$$

Taking the transpose of the second of these equations, we obtain:

$$\boldsymbol{x}_{i}^{\mathsf{T}} A \boldsymbol{x}_{i} = \boldsymbol{x}_{i}^{\mathsf{T}} \lambda_{i} \boldsymbol{x}_{i}. \tag{1.7}$$

Now the right-hand sides of equations (1.5) and (1.6) are identical. Hence $x_j^{\mathsf{T}}(\lambda_i - \lambda_j)x_i = 0$ and since $\lambda_i \neq \lambda_j$, it must be that the vectors x_i and x_j are orthogonal.