Linear Algebra and Vector Calculus Notes and Exercises

Somnath Sikdar

February 3, 2020

Contents

Chapter 1

Linear Algebra Basics

1.1 Linear Functions

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

$$L(a\boldsymbol{x} + b\boldsymbol{y}) = aL(\boldsymbol{x}) + bL(\boldsymbol{y}).$$

It follows (by induction) that for all $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_r \in \boldsymbol{R}^m$ and all $a_1, \ldots, a_r \in \boldsymbol{R}$

$$L(a_1\boldsymbol{x}_1 + \cdots + a_r\boldsymbol{x}_r) = a_1L(\boldsymbol{x}_1) + \cdots + a_rL(\boldsymbol{x}_r).$$

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

Proof. Given any vector $\boldsymbol{x} \in \boldsymbol{R}^m$, 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, $L(\boldsymbol{x}) = \sum_i \alpha_i L(e_i)$ which is completely specified by $L(e_1), \ldots, L(e_m)$.

Let $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_m$ be any vectors in $\boldsymbol{R}^n$. Define a map L from $\boldsymbol{R}^m$ to $\boldsymbol{R}^n$ as follows: for $\boldsymbol{x} = \sum_{i=1}^m \alpha_i e_i \in \boldsymbol{R}^m$, $L(\boldsymbol{x}) = \sum_{i=1}^m \alpha_i \boldsymbol{b}_i$. Then $L(e_i) = \boldsymbol{b}_i$ for all $1 \le i \le m$ and for all $\boldsymbol{x}, \boldsymbol{y} \in \boldsymbol{R}^m$ and all $a, b \in \boldsymbol{R}$:

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

$\square$

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

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

1. *$\mathrm{Im}(L)$ is a subspace of $\mathbf{R}^n$;*

2. *$\mathrm{Ker}(L)$ is a subspace of $\mathbf{R}^m$;*

3. *$\dim(\mathrm{Im}(L)) + \dim(\mathrm{Ker}(L)) = \dim(\mathbf{R}^m) = m$.*

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

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

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

To prove (3), let us assume that $\dim(\mathrm{Ker}(L)) = k$ and that $b_1, \ldots, b_k$ is a basis of $\mathrm{Ker}(L)$. Extend this basis to a basis $b_1, \ldots, b_k, b_{k+1}, \ldots, b_m$ for $\mathbf{R}^m$. Then for every vector $x \in \mathbf{R}^m$ there exist scalars $\beta_1, \ldots, \beta_m$ 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 $\mathrm{Im}(L)$ 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 $\mathrm{Im}(L)$, 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 $\sum_{i=k+1}^{m} \beta_i L(b_i) = 0_n$. By the linearity of L, we have $L(\sum_{i=k+1}^{m} \beta_i b_i) = 0_n$ and hence $0_m \neq \sum_{i=k+1}^{m} \beta_i b_i \in \mathrm{Ker}(L)$. 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 $\mathrm{Im}(L)$. This proves (3). $\square$

1.3 Quadratic Forms

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

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

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

$$(x_1, \ldots, x_m) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{m1} \\ \vdots & \vdots & \cdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mm} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix} = (x_1, \ldots, x_m)[x_1 A_{*1} + \cdots + x_m A_{*m}]$$
$$= \sum_{i=1}^{m} \sum_{j=1}^{m} x_i a_{ij} x_j.$$

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 $h^\mathsf{T}Ah$, where A is the all-ones 3×3 matrix.

1.3.1 Definiteness

A quadratic form $q \colon R^m \to R$ is

1. *positive definite* if for all non-zero $x \in R^m$, $q(x) > 0$;

2. *negative definite* if for all non-zero $x \in R^m$, $q(x) > 0$;

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 $(h_1, h_2, h_3) \in R^3 \setminus \{0\}$.

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 $\bar{x}^\mathsf{T}$ and (1.2) by x^T to obtain:

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

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

4

Take the transpose of equation (1.4), we obtain: $\bar{x}^\mathsf{T} A x = \bar{x}^\mathsf{T} \bar{\lambda} x$, 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? It turns out that the eigenvectors corresponding to the eigenvalues of a real symmetric matrix form an orthonormal basis for $\mathbf{R}^m$. We show this using a sequence of lemmas.

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_j^T and the second by x_i^T, we obtain:

$$x_j^\mathsf{T} A x_i = x_j^\mathsf{T} \lambda_i x_i \tag{1.5}$$
$$x_i^\mathsf{T} A x_j = x_i^\mathsf{T} \lambda_j x_j \tag{1.6}$$

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

$$x_j^\mathsf{T} A x_i = x_j^\mathsf{T} \lambda_j x_i. \tag{1.7}$$

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. □

Lemma 1.3. *Let $A \in \mathbf{R}^{m \times m}$ and let λ be an eigenvalue. The eigenvectors belonging to λ form a subspace $\mathscr{E}_A(\lambda)$ of $\mathbf{R}^m$.*

Proof. Let x_1 and x_2 are two eigenvectors belonging to λ. Then

$$A(a x_1 + b x_2) = a A x_1 + b A x_2 = \lambda (a x_1 + b x_2).$$

Thus any linear combination of the eigenvectors is also an eigenvector of λ, showing that this set of vectors is indeed a subspace. $\qquad\square$

Lemma 1.4. *Let $A \in \mathbf{R}^{m \times m}$ be a real symmetric matrix. Then the algebraic multiplicity of each eigenvalue of A equals its geometric multiplicity.*

Proof. Let λ_i be an eigenvalue with algebraic multiplicity k. Suppose that $\dim(\mathscr{E}_A(\lambda_i)) = r$ and let $\{v_1, \dots, v_r\}$ be an orthonormal basis for $\mathscr{E}_A(\lambda_i)$. Extend this to an orthonormal basis $\{v_1, \dots, v_r, v_{r+1}, \dots, v_m\}$ for $\mathbf{R}^m$.

Define $S = [v_1, \dots, v_r, v_{r+1}, \dots, v_m]$ to be the $m \times m$ matrix whose columns are the basis vectors. Then S is an orthonormal matrix and $S^{-1} = S^{\mathsf{T}}$. Consider the matrix $S^{-1}AS$. This is similar to A and therefore has the same eigenvalues as A including the same multiplicities.

We may write $S^{-1}AS$ as:

$$
\begin{aligned}
S^{-1}AS &= \begin{pmatrix} -v_1^{\mathsf{T}}- \\ \vdots \\ -v_m^{\mathsf{T}}- \end{pmatrix} A \begin{pmatrix} v_1 & v_2 & \cdots & v_m \\ | & | & & | \end{pmatrix} \\
&= \begin{pmatrix} -v_1^{\mathsf{T}}- \\ \vdots \\ -v_m^{\mathsf{T}}- \end{pmatrix} \begin{pmatrix} \lambda_i v_1 & \cdots & \lambda_i v_r & A v_{r+1} & \cdots & A v_m \\ | & & | & | & & | \end{pmatrix} \\
&= \left(\begin{array}{c|c} \lambda_i I_r & \mathbf{0}_{r,m-r} \\ \hline \mathbf{0}_{m-r,r} & C^{\mathsf{T}}AC \end{array} \right),
\end{aligned}
$$

where $C = [v_{r+1}, \dots, v_m]$. $\qquad\square$