Linear Algebra and Vector Calculus
Notes and Exercises

Somnath Sikdar

May 27, 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 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_r \in \boldsymbol{R}^m$ and all $a_1, \dots, a_r \in \boldsymbol{R}$

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

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 $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ of $\boldsymbol{R}^m$. An arbitrary choice of vectors $L(\boldsymbol{e}_1), \dots, L(\boldsymbol{e}_m)$ 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 $\sum_{i=1}^m \alpha_i \boldsymbol{e}_i$ of the basis vectors. By the linearity of L, $L(\boldsymbol{x}) = \sum_i \alpha_i L(\boldsymbol{e}_i)$ which is completely specified by $L(\boldsymbol{e}_1), \dots, L(\boldsymbol{e}_m)$.

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

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

□

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

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

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

Proof. The proof of (1) and (2) are similar. For proving (1), let $\boldsymbol{z}_1, \boldsymbol{z}_2 \in \mathrm{Im}(L)$ and $\alpha, \beta \in \boldsymbol{R}$. Then there exist $\boldsymbol{x}_1, \boldsymbol{x}_2 \in \boldsymbol{R}^m$ such that

$$L(\alpha \boldsymbol{x}_1 + \beta \boldsymbol{x}_2) = \alpha \boldsymbol{z}_1 + \beta \boldsymbol{z}_2,$$

implying that $\alpha \boldsymbol{z}_1 + \beta \boldsymbol{z}_2 \in \mathrm{Im}(L)$. Since $\mathrm{Im}(L)$ is closed under vector addition and multiplication by scalars, it is a vector space. Since it is a subset of $\boldsymbol{R}^n$, it must be a subspace of $\boldsymbol{R}^n$.

To prove (3), let us assume that $\dim(\mathrm{Ker}(L)) = k$ and that $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$ is a basis of $\mathrm{Ker}(L)$. Extend this basis to a basis $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k, \boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_m$ for $\boldsymbol{R}^m$. Then for every vector $\boldsymbol{x} \in \boldsymbol{R}^m$ there exist scalars $\beta_1, \ldots, \beta_m$ such that $\boldsymbol{x} = \sum_{i=1}^{k} \beta_i \boldsymbol{b}_i$. Moreover since L is linear,

$$\begin{aligned} L(\boldsymbol{x}) &= \sum_{i=1}^{m} \beta_i L(\boldsymbol{b}_i) \\ &= \sum_{i=k+1}^{m} \beta_i L(\boldsymbol{b}_i). \end{aligned}$$

This shows that every vector in $\mathrm{Im}(L)$ can be expressed as a linear combination of the vectors $L(\boldsymbol{b}_{k+1}), \ldots, L(\boldsymbol{b}_m)$. 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 $\sum_{i=k+1}^{m} \beta_i L(\boldsymbol{b}_i) = \boldsymbol{0}_n$. By the linearity of L, we have $L(\sum_{i=k+1}^{m} \beta_i \boldsymbol{b}_i) = \boldsymbol{0}_n$ and hence $\boldsymbol{0}_m \neq \sum_{i=k+1}^{m} \beta_i \boldsymbol{b}_i \in \mathrm{Ker}(L)$. This is a contradiction since the vectors $\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_m$ are not in the space spanned by the vectors $\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k$. Thus the vectors $\beta_{k+1}, \ldots, \beta_m$ must be independent and form a basis of $\mathrm{Im}(L)$. This proves (3). □

1.3 Quadratic Forms

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

$$q(\boldsymbol{x}) = \boldsymbol{x}^\intercal A \boldsymbol{x}.$$

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

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

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 $\boldsymbol{h}^\intercal A\boldsymbol{h}$, where A is the all-ones 3×3 matrix.

1.3.1 Definiteness

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

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

Example 1.2. Let $q(\boldsymbol{h}) = -h_1^2 + 2h_2^2 - h_3^2$ 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 $(h_1, h_2, h_3) \in \boldsymbol{R}^3 \setminus \{0\}$.

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 $\boldsymbol{x}$ be a eigenvector corresponding to it, which can be a complex vector. Then

$$A\boldsymbol{x} = \lambda \boldsymbol{x}. \tag{1.1}$$

Taking complex conjugates on each side, we get:

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

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

$$\bar{\boldsymbol{x}}^\intercal A\boldsymbol{x} = \bar{\boldsymbol{x}}^\intercal \lambda \boldsymbol{x} \tag{1.3}$$
$$\boldsymbol{x}^\intercal A\bar{\boldsymbol{x}} = \boldsymbol{x}^\intercal \bar{\lambda}\bar{\boldsymbol{x}}. \tag{1.4}$$

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

$$\boldsymbol{x}^\intercal(\lambda - \bar{\lambda})\bar{\boldsymbol{x}} = 0.$$

Since $\boldsymbol{x}^\intercal\bar{\boldsymbol{x}}$ is the sum of products of complex conjugates, it is not zero unless each component of $\boldsymbol{x}$ is zero. Since this is not the case ($\boldsymbol{x}$ 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 $\boldsymbol{R}^m$. 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 $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ be eigenvectors corresponding to these eigenvalues. Then $A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i$ and $A\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_j$. Pre-multiplying the first of these equations by $\boldsymbol{x}_j^\intercal$ and the second by $\boldsymbol{x}_i^\intercal$, we obtain:

$$\boldsymbol{x}_j^\intercal A\boldsymbol{x}_i = \boldsymbol{x}_j^\intercal\lambda_i\boldsymbol{x}_i \tag{1.5}$$
$$\boldsymbol{x}_i^\intercal A\boldsymbol{x}_j = \boldsymbol{x}_i^\intercal\lambda_j\boldsymbol{x}_j \tag{1.6}$$

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

$$\boldsymbol{x}_j^\intercal A\boldsymbol{x}_i = \boldsymbol{x}_j^\intercal\lambda_j\boldsymbol{x}_i. \tag{1.7}$$

Now the right-hand sides of equations (1.5) and (1.6) are identical. Hence $\boldsymbol{x}_j^\intercal(\lambda_i - \lambda_j)\boldsymbol{x}_i = 0$ and since $\lambda_i \neq \lambda_j$, it must be that the vectors $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$ are orthogonal. □

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

Proof. Let $\boldsymbol{x}_1$ and $\boldsymbol{x}_2$ are two eigenvectors belonging to λ. Then

$$A(a\boldsymbol{x}_1 + b\boldsymbol{x}_2) = aA\boldsymbol{x}_1 + bA\boldsymbol{x}_2 = \lambda(a\boldsymbol{x}_1 + b\boldsymbol{x}_2).$$

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

The algebraic multiplicity of an eigenvalue λ_i of a matrix A is the number of times λ_i appears as a root of the equation $\det(A - \lambda I) = 0$. The geometric multiplicity of λ_i is the dimension of the eigenspace $\mathscr{E}_A(\lambda)$ associated with λ_i.

Lemma 1.4. *Let $A \in \boldsymbol{R}^{m\times m}$ be a 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 $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r\}$ be an orthonormal basis for $\mathscr{E}_A(\lambda_i)$. Extend this to an orthonormal basis $\{\boldsymbol{v}_1, \dots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_m\}$ for $\boldsymbol{R}^m$.

Define $S = [\boldsymbol{v}_1, \dots, \boldsymbol{v}_r, \boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_m]$ to be the $m \times m$ matrix whose columns are the basis vectors. Then S is an orthonormal matrix and $S^{-1} = S^\intercal$. 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:

$$\begin{aligned}
S^{-1}AS &= \begin{pmatrix} -\boldsymbol{v}_1^\intercal- \\ \vdots \\ -\boldsymbol{v}_m^\intercal- \end{pmatrix} A \begin{pmatrix} \boldsymbol{v}_1 & \boldsymbol{v}_2 & \cdots & \boldsymbol{v}_m \\ | & | & & | \end{pmatrix} \\
&= \begin{pmatrix} -\boldsymbol{v}_1^\intercal- \\ \vdots \\ -\boldsymbol{v}_m^\intercal- \end{pmatrix} \begin{pmatrix} \lambda_i\boldsymbol{v}_1 & \cdots & \lambda_i\boldsymbol{v}_r & A\boldsymbol{v}_{r+1} & \cdots & A\boldsymbol{v}_m \\ | & & | & | & & | \end{pmatrix} \\
&= \left(\begin{array}{c|c} \lambda_i I_r & \mathbf{0}_{r,m-r} \\ \hline \mathbf{0}_{m-r,r} & C^\intercal AC \end{array}\right),
\end{aligned}$$

where $C = [\boldsymbol{v}_{r+1}, \dots, \boldsymbol{v}_m]$. The characteristic polynomials $\det(S^{-1}AS - \lambda I_m)$ and $\det(A - \lambda I_m)$ are identical and

$$\det(S^{-1}AS - \lambda I_m) = (\lambda_i - \lambda)^r \det(C^\intercal AC - \lambda I_{m-r}).$$

Since we know that in the RHS, the term $\lambda_i - \lambda$ is raised to the kth power, we must have $r \leq k$. $\square$

Although we do not prove it here, the fact remains that for real symmetric matrices, the algebraic multiplicity of an eigenvalue equals its geometric multiplicity. Since eigenvectors corresponding to two distinct eigenvalues of a real symmetric matrix are orthogonal (Lemma 1.2), by Lemmas 1.3 and 1.4, the eigenspaces of distinct eigenvectors are mutually orthogonal.

Lemma 1.5. *Let $A \in \boldsymbol{R}^{m\times m}$ be a symmetric matrix. Then there exist an orthonormal set of eigenvectors that form a basis for $\boldsymbol{R}^m$.*

Proof. Let $\lambda_1, \ldots, \lambda_r$ be the distinct eigenvalues of A with algebraic multiplicities $m_1, \ldots, m_r$, respectively. Then $\sum_i m_i = m$. Since the geometric multiplicity of an eigenvalue equals its algebraic multiplicity, we have $\dim(\mathscr{E}_A(\lambda_i)) = m_i$. Since the eigenspaces $\mathscr{E}_A(\lambda_1), \ldots, \mathscr{E}_A(\lambda_r)$ are mutually orthogonal, the union of their orthonormal bases is an independent set of vectors of size m. Hence this forms a basis for $\boldsymbol{R}^m$. □

We can use Lemma 1.5 to diagonalize a symmetric matrix $A \in \boldsymbol{R}^{m \times m}$. Let $\lambda_1, \ldots, \lambda_m$ be its eigenvalues, in no particular order. These eigenvalues need not be distinct. Let $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ be an orthonormal set of eigenvectors that form a basis for $\boldsymbol{R}^m$. Define $P = [\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m]$. Then $P^{-1}AP = D$, where D is a diagonal matrix whose main diagonal contains the eigenvalues $\lambda_1, \ldots, \lambda_m$.

Lemma 1.6. *Let $A \in \boldsymbol{R}^{m \times m}$ be a symmetric matrix. Then for any $\boldsymbol{x} \in R^m$,*

$$\lambda_{min} \cdot \|\boldsymbol{x}\|^2 \leq \boldsymbol{x}^\intercal A \boldsymbol{x} \leq \lambda_{max} \cdot \|\boldsymbol{x}\|^2,$$

where λ_{min} and λ_{max} are a minimum and a maximum eigenvalue of A.

Proof. Write $\boldsymbol{x} = \sum_{i=1}^m a_i \boldsymbol{u}_i$, where $\boldsymbol{u}_1, \ldots, \boldsymbol{u}_m$ are an orthonormal set of eigenvectors that form a basis for $\boldsymbol{R}^m$. Then

$$\begin{aligned}
\boldsymbol{x}^\intercal A \boldsymbol{x} &= \left(\sum_{i=1}^m a_i \boldsymbol{u}_i^\intercal\right) \cdot A \cdot \left(\sum_{i=1}^m a_i \boldsymbol{u}_i\right) \\
&= \left(\sum_{i=1}^m a_i \boldsymbol{u}_i^\intercal\right) \cdot \left(\sum_{i=1}^m a_i \lambda_i \boldsymbol{u}_i\right) \\
&= \sum_{i=1}^m a_i^2 \lambda_i \boldsymbol{u}_i^\intercal \boldsymbol{u}_i \\
&= \sum_{i=1}^m a_i^2 \lambda_i.
\end{aligned}$$

The last term lies clearly between $\lambda_{\min} \cdot \|\boldsymbol{x}\|^2$ and $\lambda_{\max} \cdot \|\boldsymbol{x}\|^2$. □

We can now state the characterization of quadractic forms.

Theorem 1.3. *A real symmetric matrix A is*

1. *positive definite iff all its eigenvalues are strictly positive;*
2. *negative definite iff all its eigenvalues are strictly negative;*
3. *indefinite iff some of its eigenvalues are positive and some are negative.*

Chapter 2

Excercises on Singular Value Decompositions

Exercise 2.1. suppose $T \in \mathscr{L}(V)$ is self-adjoint. Prove that the singular values of T are the absolute values of the eigenvalues of T repeated appropriately. This is *not* true for operators that are not self-adjoint.

Solution. Suppose that T is self-adjoint with eigenvalues $\lambda_1, \ldots, \lambda_n$. Then $T^\star T$ is positive and has eigenvalues $\bar{\lambda}_1 \lambda_1, \ldots, \bar{\lambda}_n \lambda_n$. The eigenvalues of $\sqrt{T^\star T}$ are the positive square roots of the eigenvalues of $T^\star T$. Hence the singular values of T are $|\lambda_1|, \ldots, |\lambda_n|$. ■

Exercise 2.2. Suppose $T \in \mathscr{L}(V)$. Prove that T and $T^\star$ have the same singular values.

Solution. For any $T \in \mathscr{L}(V)$, the operators $T^\star T$ and $T T^\star$ are both positive. ■