Appendix C.
Properties of Real Symmetric Matrices

A matrix A is **symmetric** if $A = A^T$ – the transpose of A. This means that $a = [a_{ij}]$ is $n \times n$ matric with $a_{ij} = a_{ji}$ for all $i, j = 1, 2, \dots, n$. We treat vector in $\mathbb{R}^n$ as column vectors: $x = (x_1, x_2, \dots, x_n)^T$, $y = (y_1, y_2, \dots, y_n)^T$, etc., with **dot** or **scalar product**

$$x \cdot y = (x, y) := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n = \sum_{k=1}^{n} x_k y_k.$$

Then

$$(Ax, y) = \sum_{i,j=1}^{n} a_{ij} x_j y_i = (x, A^T y).$$

In particular,

$$(1) \qquad (Ax, y) = (x, Ay) \quad \text{for all} \quad x, y \in \mathbb{R}^n \quad \text{if} \quad A = A^T.$$

It is easy to verify that the gradient

$$(2) \qquad \nabla (Ax, x) \equiv 2Ax \quad \text{if} \quad A = A^T.$$

Definition C.1. If $Av = \lambda v$, where $0 \neq v \in \mathbb{R}^n$, then v is an **eigenvector** of A, and λ is the corresponding eigenvalue.

The equality $Av = \lambda v$ is equivalent to $(A - \lambda I)v = 0$, where I is the unit matrix. This implies that all the eigenvalues of A are roots of the **characteristic equation**

$$(3) \qquad p_A(\lambda) := \det(A - \lambda I) = 0.$$

Theorem C.2. *For every real symmetric $n \times n$ matrix A, there is an orthonormal basis $v_1, v_2, \dots, v_n$ in $\mathbb{R}^n$ of eigenvectors of A: $Av_k = \lambda_k v_k$ for $k = 1, 2, \dots, n$, with $\lambda_k \in \mathbb{R}^1$.*

Proof. Step 1. The function (Ax, x) is continuous on the compact set $\{|x| = 1\} \subset \mathbb{R}^n$. Therefore, it attains

$$\lambda_1 := \min_{|x|=1} (Ax, x) = (Av_1, v_1) \quad \text{at some point} \quad v_1 \in \mathbb{R}^n, \ |v_1| = 1.$$

Then the function

$$f_1(x) := (Ax, x) - \lambda_1 |x|^2$$

attains its minimum value $f_1(v_1) = 0$ on $\{|x| = 1\}$. Since f_1 is homogeneous of degree 2, we have $f_1 \geq 0$ in $\mathbb{R}^n$, and $f_1(x)$ attains its local minimum at $x = v_1$. At this point, we must have, using (2):

$$\nabla f_1(x) = 2Ax - 2\lambda_1 x = 0.$$

This means $Av_1 = \lambda_1 v_1$.

Step 2. Next, consider the subspace

$$V_1 := \{x \in \mathbb{R}^n : \quad x \perp v_1, \quad \text{i.e.} \quad (x, v_1) = 0\}.$$

If $x \in V_1$, then

$$(Ax, v_1) = (x, Av_1) = (x, \lambda_1 v_1) = \lambda_1 (x, v_1) = 0,$$

i.e. $Ax \in V_1$. Therefore, $A(V_1) \subset V_1$, and we can consider A as a linear transformation of the $(n-1)$-dimensional space V_1 into itself. Since $(Ax, y) \equiv (x, Ay)$, the matrix of A in any basis of V_1 is symmetric. This is similar to the equalities

$$a_{ij} = (Ae_j, e_i) = (e_j, Ae_i) = a_{ji} \tag{4}$$

in the original basis

$$e_1 := (1, 0, \ldots, 0, 0)^T, \quad e_2 := (0, 1, \ldots, 0, 0)^T, \quad \ldots, \quad e_n := (0, 0, \ldots, 0, 1)^T \quad \text{in} \quad \mathbb{R}^n. \tag{5}$$

Therefore, the argument in Step 1 shows that

$$\lambda_2 := \min_{|x|=1,\, x \perp v_1} (Ax, x) = (Av_2, v_2), \quad \text{where} \quad Av_2 = \lambda_2 v_2,\ |v_2| = 1,\ v_2 \perp v_1.$$

Step 3. Continuing this procedure, we get the set of eigenvalues $\lambda_1 \le \lambda_2 \le \ldots \le \lambda_n$ and the orthonormal system of eigenvectors $v_1, v_2, \ldots, v_n$:

$$Av_k = \lambda_k v_k \quad \text{for all} \quad k, \quad \text{and} \quad (v_i, v_j) = \delta_{ij} := \begin{cases} 1 & \text{if} \quad j = k, \\ 0 & \text{if} \quad j \ne k. \end{cases} \tag{6}$$

□

Lemma C.3. *If* $(v_i, v_j) = \delta_{ij}$ *for all* $i, j = 1, 2, \ldots, n$, *then the matrix*

$$S := [v_1,\ v_2,\ \cdots, v_n] \quad \text{with columns} \quad v_1,\ v_2,\ \cdots, v_n \tag{7}$$

is **orthogonal**, *i.e.* $S^{-1} = S^T$. *In the new coordinates* $y_1, y_2, \ldots, y_n$ *with respect to the orthonormal basis* $v_1,\ v_2,\ \cdots, v_n$, *which satisfies (6), we have*

$$(Ax, x) = \sum_{i,j=1}^{n} a_{ij} x_i x_j = \sum_{k=1}^{n} \lambda_k y_k^2. \tag{8}$$

Proof. The matric $C = [c_{ij}] := S^T S$ has entries

$$\begin{aligned} c_{ij} &= (i^{th} \text{ row of } S^T) \cdot (j^{th} \text{ column of } S) \\ &= (i^{th} \text{ column of } S) \cdot (j^{th} \text{ column of } S) \\ &= (v_i, v_j) = \delta_{ij}. \end{aligned}$$

This means that $C := S^T S = I$ – the unit matrix, and $S^{-1} = S^T$.

In order to verify the equality (8), let $B = [b_{ij}]$ be the matrix of the transformation A in the basis $v_1,\ v_2,\ \cdots, v_n$. Then similarly to (4), we have

$$b_{ij} = (v_j, Av_i) = (v_j, \lambda_i v_i) = \lambda_i\,(v_j, v_i) = \lambda_i \delta_{ij},$$

hence

$$(Ax, x) = (By, y) = \sum_{i,j=1}^{n} b_{ij} y_i y_j = \sum_{k=1}^{n} \lambda_k y_k^2.$$

□

Remark C.4. If $v_1, v_2, \cdots, v_n$ is a basis in $\mathbb{R}^n$, i.e. linearly independent eigenvectors of $n \times n$ matrix $A = [a_{ij}]$, which is not necessarily symmetric, then $Av_k = \lambda_k v_k$ for $k = 1, 2, \ldots, n$, without the orthogonality condition $(v_i, v_j) = \delta_{ij}$. In this case, the matrices A and S in (7) still satisfy

$$\begin{aligned} AS &= [Av_1, Av_2, \cdots, Av_n] = [\lambda_1 v_1, \lambda_2 v_2, \cdots, \lambda_n v_n] \\ &= [v_1, v_2, \cdots, v_n] \cdot \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 & 0 \\ 0 & \lambda_2 & \cdots & 0 & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & 0 & \lambda_n \end{bmatrix} = S\Lambda, \end{aligned}$$

where $\Lambda := \operatorname{diag}[\lambda_1, \lambda_1, \ldots, \lambda_n]$. This implies

$$A = S\Lambda S^{-1}, \tag{9}$$

i.e. A is **similar** to the diagonal matrix Λ. Note that the characteristic polynomials (3) for similar matrices coincide: if $A = SBS^{-1}$, then

$$\begin{aligned} p_A(\lambda) &:= \det(A - \lambda I) = \det(SBS^{-1} - S \cdot \lambda I \cdot S^{-1}) = \det(S \cdot (B - \lambda I) \cdot S^{-1}) \\ &= \det S \cdot \det(B - \lambda I) \cdot \det(S^{-1}) = \det(B - \lambda I) = p_B(\lambda). \end{aligned}$$

In our case $B = \Lambda$, from (9) it follows

$$p_A(\lambda) = p_\Lambda(\lambda) = \prod_{k=1}^{n} (\lambda_k - \lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \cdots (\lambda_n - \lambda). \tag{10}$$

Definition C.5. The **trace** of a square matrix $A = [a_{ij}]$ is the sum of its diagonal elements:

$$\operatorname{tr} A = \operatorname{tr}[a_{ij}] := \sum_{i=1}^{n} a_{ii}.$$

Lemma C.6. *If A is a $m \times n$ matrix, and B is a $n \times m$ matrix, then the $m \times m$ matrix AB and the $n \times n$ matrix BA have same trace:* $\operatorname{tr}(AB) = \operatorname{tr}(BA)$.

Proof. If $C = [c_{ij}] = AB$, then

$$c_{ij} = (i^{th} \text{ row of } A) \cdot (j^{th} \text{ column of } B) = \sum_{k=1}^{n} a_{ik} b_{kj},$$

and

$$\operatorname{tr}(AB) = \operatorname{tr} C = \sum_{i=1}^{m} c_{ii} = \sum_{i,k} a_{ik} b_{ki}.$$

Since the last expression is symmetric with respect to A and B, we get $\operatorname{tr}(AB) = \operatorname{tr}(BA)$. □

Lemma C.7. *If $n \times n$ matrix A has n linearly independent eigenvectors $v_1, v_2, \cdots, v_n$, i.e. $Av_k = \lambda_k v_k$ for $k = 1, 2, \ldots, n$, then*

$$\det A = \prod_{k=1}^{n} \lambda_k = \lambda_1 \lambda_2 \cdots \lambda_n, \qquad \operatorname{tr} A = \sum_{k=1}^{n} \lambda_k = \lambda_1 + \lambda_2 + \cdots + \lambda_n. \tag{11}$$

Proof. The first equality in (11) follows from (10) with $\lambda = 0$. One can also get the second equality in (11) by comparing the coefficients of λ in both sides of (10).

Alternatively, one can apply Lemma C.6 to (9) as follows:

$$\operatorname{tr} A = \operatorname{tr}(S \cdot \Lambda S^{-1}) = \operatorname{tr}(\Lambda S^{-1} \cdot S) = \operatorname{tr} \Lambda = \sum_{k=1}^{n} \lambda_k.$$

□

Theorem C.8. *Let A be a real symmetric $n \times n$ matrix with eigenvalues $\lambda_1, \lambda_2, \ldots, \lambda_n$, i.e according to (10),*

$$p_A(\lambda) := \det(A - \lambda I) = \prod_{k=1}^{n} (\lambda_k - \lambda).$$

Then $f(x) := (Ax, x)$ satisfies $\nabla f(0) = 0$. In addition,

(i) *if there are λ_k of different sign: $\lambda_{k_1} < 0 < \lambda_{k_2}$, then $f(x)$ has neither maximum nor minimum at $x = 0$;*

(ii) *if $\lambda_k < 0$ for all k, then $f(x)$ has a local maximum at $x = 0$;*

(iii) *if $\lambda_k > 0$ for all k, then $f(x)$ has a local minimum at $x = 0$.*

Proof. By (2), we have $\nabla f(x) = 2Ax$, so that $\nabla f(0) = 0$. The properties (i)–(iii) follow directly from the representation of $f(x) := (Ax, x)$ in (8). □

Corollary C.9. *In the case $n = 2$, the conditions (i)–(iii) in the previous theorem are simplified as follows:*

(i) *if $\det A < 0$, then $f(x) := (Ax, x)$ has neither maximum nor minimum at $x = 0$;*

(ii) *if $\det A > 0$ and $\operatorname{tr} A < 0$, then $f(x)$ has a local maximum at $x = 0$;*

(iii) *if $\det A > 0$ and $\operatorname{tr} A > 0$, then $f(x)$ has a local minimum at $x = 0$.*

Proof. In the case $n = 2$, the equalities (11) have the form $\det A = \lambda_1 \lambda_2$ and $\operatorname{tr} A = \lambda_1 + \lambda_2$. We have $\det A < 0$ if and only if λ_1 and λ_2 have opposite signs, and $\det A > 0$ if and only if λ_1 and λ_2 have same sign. Hence the properties (i)–(iii) in this corollary follow from the corresponding properties in Theorem C.8. □