# 7 Diagonalization and Quadratic Forms

## Diagonalization

Recall the definition of a **diagonal** matrix from Section 1.6.

**Definition 7.1.** *A square matrix $A$ is* ***diagonalizable*** *if there exists an invertible matrix $P$ such that $P^{-1}AP$ is diagonal. We say that $P$* ***diagonalizes*** *$A$.*

**Remark.** Why is this interesting? For many applications we need to compute **powers** of matrices, for example

$$A^7 = AAAAAAA.$$

To do this by direct calculation is a lot of work, but if $A$ is diagonalizable, say $P^{-1}AP = D$ diagonal, then $A = PDP^{-1}$ so

$$A^7 = (PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1})(PDP^{-1}) = PD^7P^{-1}.$$

and more generally, $A^k = PD^kP^{-1}$ for all $k$. We have seen in Exercise Sheet 5 that $D^k$ is easy to compute, so this gives a much easier way to work out $A^k$ for large $k$.

**Examples 7.2.** Consider

$$A = \begin{pmatrix} 0 & 0 & -2 \\ 1 & 2 & 1 \\ 1 & 0 & 3 \end{pmatrix}, P = \begin{pmatrix} -1 & 0 & -2 \\ 0 & 1 & 1 \\ 1 & 0 & 1 \end{pmatrix}, P^{-1} = \begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ -1 & 0 & -1 \end{pmatrix}.$$

Then we can check that

$$P^{-1}AP = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix} =: D,$$

so $A$ is diagonalizable. Now for any $m$ we have

$$A^m = PD^mP^{-1} = P \begin{pmatrix} 2^m & 0 & 0 \\ 0 & 2^m & 0 \\ 0 & 0 & 1^m \end{pmatrix} P^{-1}$$

so to calculate $A^m$ we need only calculate the scalar power $2^m$ and then perform two matrix multiplications.

**Exercise.** With $A$ as above, work out $A^{16}$. Then try and do it directly without the "diagonalization"!

Given that diagonalizing a matrix is so useful, it is natural to ask which matrices can be diagonalized. To answer this question we will need a lemma giving yet another characterisation of invertible matrices.

**Lemma 7.3.** *Let $P$ be an $n \times n$ square matrix. Then $P$ is invertible if and only if its columns (viewed as column $n$-vectors) form a set of $n$ linearly independent vectors.*

*Proof.* See Section 14. $\square$

**Theorem 7.4.** *Let $A$ be an $n \times n$ matrix. Then $A$ is diagonalizable if and only if $A$ has $n$ linearly independent eigenvectors. A matrix $P$ diagonalizes $A$ if and only if $P$'s columns form a set of $n$ linearly independent eigenvectors for $A$. If it does, then the main diagonal entries of the diagonal matrix $P^{-1}AP$ are the eigenvalues of $A$ (in the order corresponding to the columns of $P$).*

*Proof.* Suppose $P^{-1}AP = D$ is diagonal. Let $c_1, \dots, c_n$ be the columns of $P$. By Lemma 7.3, the columns are linearly independent. Now $P^{-1}AP = D$ implies

$$AP = PD.$$

It follows from the definition of matrix multiplication that (i) the $i$th column of $PD$ is $D_{ii}c_i$ and (ii) the $i$th column of $AP = Ac_i$. Thus we have $Ac_i = D_{ii}c_i$, so each column $c_i$ is an eigenvector of $A$ corresponding to the eigenvalue $D_{ii}$.

Conversely, if $A$ has $n$ linearly independent eigenvectors $c_1, \dots, c_n$ then let $P$ be the matrix with these as columns, and $D$ the diagonal matrix with the corresponding eigenvalues on the main diagonal. By Lemma 7.3, $P$ is invertible. Now reversing the argument above, the $i$th column of $AP$ is $Ac_i$ and the $i$th column of $PD$ is $D_{ii}c_i$ so $AP = PD$, so $P^{-1}AP = D$. $\square$

**Example 7.5.** Let

$$A = \begin{pmatrix} 3 & 0 \\ 8 & -1 \end{pmatrix}, u = \begin{pmatrix} 1 \\ 2 \end{pmatrix} \text{ and } v = \begin{pmatrix} 0 \\ 1 \end{pmatrix}.$$

Then $A$ has eigenvalues 3 and $-1$, with corresponding eigenvectors $u$ and $v$ respectively (**exercise: check this**). It is easy to check that $\{u, v\}$ is linearly independent. If we let $P$ be the matrix whose columns are $u$ and $v$,

$$P = \begin{pmatrix} 1 & 0 \\ 2 & 1 \end{pmatrix}, \quad \text{then} \quad P^{-1} = \begin{pmatrix} 1 & 0 \\ -2 & 1 \end{pmatrix}$$

and one can check (**exercise**) that

$$P^{-1}AP = \begin{pmatrix} 3 & 0 \\ 0 & -1 \end{pmatrix}.$$

**Definition 7.6.** *An $n \times n$ matrix $A$ is called **orthogonally diagonalizable** if there is an **orthogonal** matrix $P$ such that $P^{-1}AP = P^TAP$ is diagonal.*

**Theorem 7.7.** *Let $A$ be an $n \times n$ matrix. Then the following are equivalent:*

*(i) $A$ is orthogonally diagonalizable.*

*(ii) $A$ has an orthonormal set of $n$ eigenvectors;*

*(iii) $A$ is symmetric.*

*Proof.* $(i) \Leftrightarrow (ii)$ This follows from Theorems 6.6, 6.8 and 7.4 (**exercise: write down exactly how**).

$(i) \Rightarrow (iii)$ If (i) holds, say $P^{-1}AP = D$ is diagonal with $P$ orthogonal, then we have $A = PDP^{-1} = PDP^T$. Clearly $D$ is symmetric, so

$$A^T = (PDP^T)^T = (P^T)^T D^T P^T = PDP^T = A$$

which means that $A$ is symmetric.

$(iii) \Rightarrow (ii)$ Omitted (see for example the textbook of Anton).

$\square$

## Quadratic Forms

**Definition 7.8.** *A quadratic form in $n$ variables is a function $f : \mathbb{R}^n \to \mathbb{R}$ of the form*

$$f(x) = f(x_1, ..., x_n) = \sum_{1 \le i \le j \le n} c_{ij} x_i x_j \qquad (*)$$

*where $x \in \mathbb{R}^n$ and $c_{ij} \in \mathbb{R}(1 \le i \le j \le n)$. Alternatively, a quadratic form is a homogeneous polynomial of degree 2 in $n$ variables $x_1, ..., x_n$.*

**Examples 7.9.** The following are quadratic forms:

1. $f(x_1) = x_1^2$
2. $f(x_1, x_2) = 2x_1^2 + 3x_2^2 - x_1x_2$
3. $f(x_1, x_2, x_3) = x_1^2 + x_2^2 + x_3^2 - 2x_1x_2 - 2x_1x_3 - 2x_2x_3$

4. $f(x_1, \ldots, x_n) = x_1^2 + x_2^2 + \cdots + x_n^2 = \langle v|v\rangle$, where $v = (x_1, \ldots, x_n)$. So the Euclidean inner product (see Chapter 6) gives rise to a quadratic form.

If we set $a_{ii} = c_{ii}$ for $i = 1, ..., n$ and $a_{ij} = \frac{1}{2}c_{ij}$ for $1 \leq i < j \leq n$, then $(*)$ becomes

$$f(x) = \sum_{i=1}^{n} a_{ii}x_i^2 + \sum_{1\leq i<j\leq n} 2a_{ij}x_ix_j$$

and we can write this as

$$f(x) = x^TAx$$

where $A$ is the symmetric $n\times n$ matrix with $(i,j)$-th entry equal to $a_{ij}$. Then $A$ is called the matrix of the quadratic form $f$.

**Example 7.10.** Let $f(x_1, x_2) = 2x_1^2 - 3x_2^2 - x_1x_2$. Then

$$f(x_1, x_2) = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 2 & -1/2 \\ -1/2 & -3 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}.$$

Now that a symmetric matrix is involved, we can take advantage of Theorem 7.7.

ie. there exists an orthogonal matrix $Q$ such that

$$Q^TAQ = D = \begin{pmatrix} \lambda_1 & ... & 0 \\ \vdots & \ddots & \vdots \\ 0 & ... & \lambda_n \end{pmatrix}$$

where $D$ is a diagonal matrix and $\lambda_1, ..., \lambda_n$ are the eigenvalues of $A$.

Now let $y = Q^{-1}x = Q^Tx$. Then $x = Qy$ and

$$f(x) = (Qy)^TA(Qy) = y^TQ^TAQy = y^TDy$$

and if $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}$ then

$$y^TDy = \lambda_1y_1^2 + ... + \lambda_ny_n^2,$$

a quadratic form in variables $y_1, ..., y_n$ with no cross terms. This process is called **diagonalization** of the quadratic form $f$. We have just proved a famous theorem, namely

**Theorem 7.11.** *(The Principal Axes Theorem) Every quadratic form $f$ can be diagonalized. More specifically, if $f(x) = x^T Ax$ is a quadratic form in* $x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$*, then there exists an orthogonal matrix $Q$ such that*

$$f(x) = x^T Ax = \lambda_1 y_1^2 + ... + \lambda_n y_n^2$$

*where* $\begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = Q^T \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}$ *and $\lambda_1, ..., \lambda_n$ are the eigenvalues of the matrix $A$.*

From the first part of the course we know that $Q$ is the matrix whose columns are the **unit** eigenvectors of the matrix $A$ of $f$.

**Example 7.12.** Let $f(x_1, x_2, x_3) = 2x_1x_2 + 2x_1x_3 + 2x_2x_3$. The matrix of $f$ is

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

The eigenvalues of $A$ are $2, -1, -1$ with corresponding unit eigenvectors

$$\begin{pmatrix} 1/\sqrt{3} \\ 1/\sqrt{3} \\ 1/\sqrt{3} \end{pmatrix}, \begin{pmatrix} -2/\sqrt{6} \\ 1/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}, \begin{pmatrix} 0 \\ 1/\sqrt{2} \\ -1/\sqrt{2} \end{pmatrix}$$

and so we have

$$Q = \begin{pmatrix} 1/\sqrt{3} & -2/\sqrt{6} & 0 \\ 1/\sqrt{3} & 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{3} & 1/\sqrt{6} & -1/\sqrt{2} \end{pmatrix}.$$

If we set $y = Q^T x$, we get

$$y_1 = \frac{1}{\sqrt{3}}(x_1 + x_2 + x_3), y_2 = \frac{1}{\sqrt{6}}(-2x_1 + x_2 + x_3), y_3 = \frac{1}{\sqrt{2}}(x_2 - x_3).$$

Then, expressed in terms of the variables $y_1, y_2$ and $y_3$, the quadratic form becomes $2y_1^2 - y_2^2 - y_3^2$.

**Definition 7.13.** *A quadratic form $f : \mathbb{R}^n \to \mathbb{R}$ is **positive definite** if $f(x) > 0$ for all $x \neq \underline{0}$.*

An immediate consequence of the Principal Axes Theorem is the following:

**Theorem 7.14.** *Let $f(x) = x^T Ax$ be a quadratic form with matrix $A$. Then $f$ is positive definite if and only if all the eigenvalues of $A$ are positive.*

*Proof.* By the Principal Axes Theorem, there exists an orthogonal matrix $Q$ such that

$$f(x) = \lambda_1 y_1^2 + ... + \lambda_n y_n^2$$

where $y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = Q^T x$ and $\lambda_1, ..., \lambda_n$ are the eigenvalues of $A$. If all the $\lambda_i$ are positive then $f(x) > 0$ except when $y = \underline{0}$. But this happens if and only if $x = \underline{0}$ because $Q^T$ is invertible. Therefore $f$ is positive definite.

On the other hand if one of the eigenvalues $\lambda_i \leq 0$, letting $y = e_i$ and $x = Qy$ we get $f(x) = \lambda_i \leq 0$ and so $f$ is not positive definite. □

We say that a symmetric matrix $A$ is **positive definite** if the associated quadratic form

$$f(x) = x^T Ax$$

is positive definite.

The Principal Axes Theorem has important applications in geometry.