Lecture Notes: Orthogonal and Symmetric Matrices

Yufei Tao
Department of Computer Science and Engineering
Chinese University of Hong Kong
taoyf@cse.cuhk.edu.hk

1 Orthogonal Matrix

Definition 1. Let $u = [u_{i1}]$ and $v = [v_{i1}]$ be two $n \times 1$ vectors. Define the dot product between them — denoted as $u \cdot v$ — as the real value $\sum_{i=1}^{n} u_{i1}v_{i1}$.

Similarly, let $u = [u_{1j}]$ and $v = [v_{1j}]$ be two $1 \times n$ vectors. Define the dot product between them — again denoted as $u \cdot v$ — as the real value $\sum_{j=1}^{n} u_{1j}v_{1j}$.

Definition 2. Let $u$ be a vector. Define $|u| = \sqrt{u \cdot u}$, and call it the norm or the length of $u$.

Definition 3. Let $S$ be a set of non-zero (column or row) vectors $v_1, v_2, ..., v_k$ of the same dimensionality. We say that $S$ is orthogonal if $v_i \cdot v_j = 0$ for any $i \neq j$. Furthermore, we say that $S$ is orthonormal if (i) $S$ is orthogonal, and (ii) $|v_i| = 1$ for any $i \in [1,k]$.

For example,

$$\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

is orthogonal but not orthonormal. If, however, we scale each of the above vectors to have length 1, then the resulting vector set becomes orthonormal:

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

Lemma 1. An orthogonal set of vectors must be linearly independent.

Proof. Suppose that $S = \{v_1, v_2, ..., v_k\}$. Assume, on the contrary, that $S$ is not linearly independent. Hence, there exist real values $c_1, c_2, ..., c_k$ that are not all zero, and make the following hold:

$$c_1v_1 + c_2v_2 + ... + c_kv_k = 0.$$

Suppose, without loss of generality, that $c_i \neq 0$ for some $i \in [1,k]$. Then, we multiply both sides of the above equation by $v_i$, and obtain:

$$c_1v_1 \cdot v_i + c_2v_2 \cdot v_i + ... + c_kv_k \cdot v_i = 0 \Rightarrow c_iv_i \cdot v_i = 0.$$

The above equation contradicts the fact that $c_i \neq 0$ and $v_i$ is a non-zero vector. □
Definition 4. An \( n \times n \) matrix \( A \) is orthogonal if (i) its inverse \( A^{-1} \) exists, and (ii) \( A^T = A^{-1} \).

Example 1. Consider \( A = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \). It is orthogonal because \( A^T = A^{-1} = \begin{bmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{bmatrix} \). The following is a 3 \( \times \) 3 orthogonal matrix:

\[
\begin{bmatrix}
\frac{2}{3} & \frac{1}{3} & \frac{2}{3} \\
-\frac{2}{3} & \frac{2}{3} & \frac{1}{3} \\
\frac{1}{3} & \frac{2}{3} & -\frac{2}{3}
\end{bmatrix}
\]

An orthogonal matrix must be formed by an orthonormal set of vectors:

Lemma 2. Let \( A \) be an \( n \times n \) matrix with row vectors \( r_1, r_2, ..., r_n \), and column vectors \( c_1, c_2, ..., c_n \). Both the following statements are true:

- \( A \) is orthogonal if and only if \( \{r_1, r_2, ..., r_n\} \) is orthonormal.
- \( A \) is orthogonal if and only if \( \{c_1, c_2, ..., c_n\} \) is orthonormal.

Proof. We will prove only the first statement because applying the same argument on \( A^T \) proves the second. Let \( B = A A^T \). Denote by \( b_{ij} \) the element of \( B \) at the \( i \)-th row and \( j \)-th column. We know that \( b_{ij} = r_i \cdot r_j \) (note that the \( j \)-th column of \( A^T \) has the same components as \( r_j \)). \( A \) is orthogonal if and only if \( B \) is an identity matrix, which in turn is true if and only if \( b_{ij} = 1 \) when \( i = j \), and \( b_{ij} = 0 \) otherwise. The lemma thus follows.

2 Symmetric Matrix

Recall that an \( n \times n \) matrix \( A \) is symmetric if \( A = A^T \). In this section, we will learn several nice properties of such matrices.

Lemma 3. All the eigenvalues of a symmetric matrix must be real values (i.e., they cannot be complex numbers). All eigenvectors of the matrix must contain only real values.

We omit the proof of the lemma (which is not difficult, but requires the definition of matrices on complex numbers). Note that the above lemma is not true for general square matrices (i.e., it is possible for an eigenvalue to be a complex number).

Lemma 4. Let \( \lambda_1 \) and \( \lambda_2 \) be two different eigenvalues of a symmetric matrix \( A \). Also, suppose that \( x_1 \) is an eigenvector of \( A \) corresponding to \( \lambda_1 \), and \( x_2 \) is an eigenvector of \( A \) corresponding to \( \lambda_2 \). It must holds that \( x_1 \cdot x_2 = 0 \).

Proof. By definition of eigenvalue and eigenvector, we know:

\[
\begin{align*}
Ax_1 &= \lambda_1 x_1 \\
Ax_2 &= \lambda_2 x_2
\end{align*}
\]
From (1), we have
\[ x_1^T A^T = \lambda_1 x_1^T \Rightarrow \]
\[ x_1^T A = \lambda_1 x_1 \Rightarrow \]
\[ x_1^T A x_2 = \lambda_1 x_1^T x_2 \Rightarrow \text{(by (2))} \]
\[ x_1^T \lambda_2 x_2 = \lambda_1 x_1^T x_2 \Rightarrow \]
\[ x_1^T x_2 (\lambda_1 - \lambda_2) = 0 \Rightarrow \text{(by \( \lambda_1 \neq \lambda_2 \))} \]
\[ x_1^T x_2 = 0. \]

The lemma then follows from the fact that \( x_1 \cdot x_2 = x_1^T x_2 \).

\[ \square \]

Example 2. Consider

\[ A = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix} \]

We know that \( A \) has two eigenvalues \( \lambda_1 = 1 \) and \( \lambda_2 = -2 \).

For eigenvalue \( \lambda_1 = 1 \), all the eigenvectors can be represented as \( x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \) satisfying:

\[ x_1 = v - u, x_2 = u, x_3 = v \]

with \( u, v \in \mathbb{R} \). Setting \((u, v)\) to \((1, 0)\) and \((0, 1)\) respectively gives us two linearly independent eigenvectors:

\[ x_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, x_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix} \]

For eigenvalue \( \lambda_2 = -2 \), all the eigenvectors can be represented as \( x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \) satisfying:

\[ x_1 = -t, x_2 = -t, x_3 = t \]

with \( t \in \mathbb{R} \). Setting \( t = 1 \) gives us another eigenvector:

\[ x_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix} \]

Vectors \( x_1, x_2, \) and \( x_3 \) are linearly independent. According to Lemma 4, both \( x_1 \cdot x_3 \) and \( x_2 \cdot x_3 \) must be 0, which is indeed the case.

\[ \square \]

From an earlier lecture, we already know that every symmetric matrix can be diagonalized because it definitely has \( n \) linearly independent eigenvectors. The next lemma strengthens this fact:
Lemma 5. Every $n \times n$ symmetric matrix has an orthogonal set of $n$ eigenvectors.

We omit the proof of the lemma (which is rather non-trivial). Note that $n$ eigenvectors in the lemma must be linearly independent, according to Lemma 1.

Example 3. Let us consider again the matrix $A$ in Example 2. We have obtained eigenvectors $x_1, x_2, x_3$. Clearly, they do not constitute an orthogonal set because $x_1, x_2$ are not orthogonal. We will replace $x_2$ with a different $x_2'$ that is still an eigenvector of $A$ for eigenvalue $\lambda_1 = 1$, and is orthogonal to $x_1$.

From Example 2, we know that all eigenvectors corresponding to $\lambda_1$ have the form $\begin{bmatrix} v - u \\ u \\ v \end{bmatrix}$. For such a vector to be orthogonal to $x_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$, we need:

\[
(-1)(v - u) + u = 0 \Rightarrow v = 2u
\]

As you can see, there are infinitely many such vectors, any of which can be $x_2'$ except $\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$. To produce one, we can choose $u = 1, v = 2$, which gives $x_2' = \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}$.

$\{x_1, x_2', x_3\}$ is thus an orthogonal set of eigenvectors of $A$.

Corollary 1. Every $n \times n$ symmetric matrix has an orthonormal set of $n$ eigenvectors.

Proof. The orthonormal set can be obtained by scaling all vectors in the orthogonal set of Lemma 5 to have length 1.

Now we prove an important lemma about symmetric matrices.

Lemma 6. Let $A$ be an $n \times n$ symmetric matrix. There exist an orthogonal matrix $Q$ such that $A = Q \, \text{diag}[\lambda_1, \lambda_2, ..., \lambda_n] \, Q^{-1}$, where $\lambda_1, \lambda_2, ..., \lambda_n$ are eigenvalues of $A$.

Proof. From an earlier lecture, we know that given a set of linearly independent eigenvectors $v_1, v_2, ..., v_n$ corresponding to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ respectively, we can produce $Q$ by placing $v_i$ as the $i$-th column of $Q$, for each $i \in [1, n]$, such that $A = Q \, \text{diag}[\lambda_1, \lambda_2, ..., \lambda_n] \, Q^{-1}$. From Corollary 1, we know that we can find an orthonormal set of $v_1, v_2, ..., v_n$. By Lemma 2, it follows that $Q$ is an orthogonal matrix.

Example 4. Consider once again the matrix $A$ in Example 2. In Example 3, we have obtained an orthogonal set of eigenvectors:

\[
\begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}
\]
By scaling, we obtain the following orthonormal set of eigenvectors:

\[
\begin{bmatrix}
-1/\sqrt{2} \\
1/\sqrt{2} \\
0
\end{bmatrix}, \begin{bmatrix}
1/\sqrt{6} \\
1/\sqrt{6} \\
2/\sqrt{6}
\end{bmatrix}, \begin{bmatrix}
-1/\sqrt{3} \\
-1/\sqrt{3} \\
1/\sqrt{3}
\end{bmatrix}
\]

Recall that these eigenvectors correspond to eigenvalues 1, 1, and –2, respectively. We thus produce:

\[
Q = \begin{bmatrix}
-1/\sqrt{2} & 1/\sqrt{6} & -1/\sqrt{3} \\
1/\sqrt{2} & 1/\sqrt{6} & 1/\sqrt{3} \\
0 & 2/\sqrt{6} & 1/\sqrt{3}
\end{bmatrix}
\]

such that \( A = Q \text{diag}[1, 1, -2] Q^{-1} \).