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. *An $n \times n$ matrix $\boldsymbol{A}$ is* **orthogonal** *if (i) its inverse $\boldsymbol{A}^{-1}$ exists, and (ii) $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$.*

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

$$\begin{bmatrix} 2/3 & 1/3 & 2/3 \\ -2/3 & 2/3 & 1/3 \\ 1/3 & 2/3 & -2/3 \end{bmatrix}$$

$\square$

Lemma 1. *If $\boldsymbol{A}$ is orthogonal, then $\boldsymbol{A}^T$ is also orthogonal.*

Proof.

$$(\boldsymbol{A}^T)^T = (\boldsymbol{A}^{-1})^T = (\boldsymbol{A}^T)^{-1}$$

The lemma thus follows. $\square$

To explain the next property of orthogonal matrices, we need to define two new concepts. Let S be a set of non-zero vectors $\boldsymbol{v}_1$, $\boldsymbol{v}_2$, ..., $\boldsymbol{v}_k$ of the same dimensionality. We say that S is *orthogonal* if $\boldsymbol{v}_i \cdot \boldsymbol{v}_j = 0$ for any $i \neq j$. Furthermore, we say that S is *orthonormal* if (i) S is orthogonal, and (ii) $|\boldsymbol{v}_i| = \boldsymbol{v}_i \cdot \boldsymbol{v}_i = 1$ for any $i \in [1, k]$. For example,

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

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:

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

Lemma 2. *An orthogonal set of vectors must be linearly independent.*

Proof. Suppose that $S = \{\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_k\}$. 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:

$$c_1\boldsymbol{v}_1 + c_2\boldsymbol{v}_2 + ... + c_k\boldsymbol{v}_k \quad = \quad \mathbf{0}.$$

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 $\boldsymbol{v}_i$, and obtain:

$$\begin{aligned} c_1\boldsymbol{v}_1 \cdot \boldsymbol{v}_i + c_2\boldsymbol{v}_2 \cdot \boldsymbol{v}_i + ... + c_k\boldsymbol{v}_k \cdot \boldsymbol{v}_i &= \mathbf{0} \;\Rightarrow \\ c_i\boldsymbol{v}_i \cdot \boldsymbol{v}_i &= \mathbf{0}. \end{aligned}$$

The above equation contradicts the fact that $c_i \neq 0$ and $\boldsymbol{v}_i$ is a non-zero vector. □

We are now ready to reveal another way to define orthogonal matrix:

Lemma 3. *Let $\boldsymbol{A}$ be an $n \times n$ matrix with row vectors $\boldsymbol{r}_1$, $\boldsymbol{r}_2$, ..., $\boldsymbol{r}_n$, and column vectors $\boldsymbol{c}_1$, $\boldsymbol{c}_2$, ..., $\boldsymbol{c}_n$. Both the following statements are true:*

- *$\boldsymbol{A}$ is orthogonal if and only if $\{\boldsymbol{r}_1, \boldsymbol{r}_2, ..., \boldsymbol{r}_n\}$ is orthonormal.*
- *$\boldsymbol{A}$ is orthogonal if and only if $\{\boldsymbol{c}_1, \boldsymbol{c}_2, ..., \boldsymbol{c}_n\}$ is orthonormal.*

Proof. We will prove only the first statement because applying the same argument on $\boldsymbol{A}^T$ proves the second. Let $\boldsymbol{B} = \boldsymbol{A}\boldsymbol{A}^T$. Denote by b_{ij} the element of B at the i-th row and j-th column. We know that $b_{ij} = \boldsymbol{r}_i \cdot \boldsymbol{r}_j$ (note that the j-th column of $\boldsymbol{A}^T$ has the same components as $\boldsymbol{r}_j$). $\boldsymbol{A}$ is orthogonal if and only if $\boldsymbol{B}$ 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. □

Lemma 4. *The determinant of an orthogonal matrix $\boldsymbol{A}$ can only be 1 or -1.*

Proof. From $\boldsymbol{A}^T = \boldsymbol{A}^{-1}$, we know that $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}$ where $\boldsymbol{I}$ is an identity matrix. Hence, $det(\boldsymbol{A}\boldsymbol{A}^T) = det(\boldsymbol{A})det(\boldsymbol{A}^T) = (det(\boldsymbol{A}))^2 = 1$. The lemma thus follows. □

2 Symmetric Matrix

Recall that an $n \times n$ matrix $\boldsymbol{A}$ is symmetric if $\boldsymbol{A} = \boldsymbol{A}^T$. Next, we give several nice properties of such matrices.

Lemma 5. *All the eigenvalues of a symmetric matrix must be real values (i.e., they cannot be complex numbers).*

We omit the proof of the lemma. 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 6. *Let λ_1 and λ_2 be two different eigenvalues of a symmetric matrix $\boldsymbol{A}$. Also, suppose that $\boldsymbol{x_1}$ is an eigenvector of $\boldsymbol{A}$ corresponding to λ_1, and $\boldsymbol{x_2}$ is an eigenvector of $\boldsymbol{A}$ corresponding to λ_2. It must holds that $\boldsymbol{x_1} \cdot \boldsymbol{x_2} = 0$.*

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

$$\begin{aligned} \boldsymbol{A}\boldsymbol{x_1} &= \lambda_1 \boldsymbol{x_1} && (1) \\ \boldsymbol{A}\boldsymbol{x_2} &= \lambda_2 \boldsymbol{x_2} && (2) \end{aligned}$$

From (1), we have

$$\begin{aligned} \boldsymbol{x_1}^T \boldsymbol{A}^T &= \lambda_1 \boldsymbol{x_1}^T \Rightarrow \\ \boldsymbol{x_1}^T \boldsymbol{A} &= \lambda_1 \boldsymbol{x_1}^T \Rightarrow \\ \boldsymbol{x_1}^T \boldsymbol{A}\boldsymbol{x_2} &= \lambda_1 \boldsymbol{x_1}^T \boldsymbol{x_2} \Rightarrow \text{(by (2))} \\ \boldsymbol{x_1}^T \lambda_2 \boldsymbol{x_2} &= \lambda_1 \boldsymbol{x_1}^T \boldsymbol{x_2} \Rightarrow \\ \boldsymbol{x_1}^T \boldsymbol{x_2}(\lambda_1 - \lambda_2) &= 0 \Rightarrow \text{(by } \lambda_1 \neq \lambda_2\text{)} \\ \boldsymbol{x_1}^T \boldsymbol{x_2} &= 0. \end{aligned}$$

The lemma then follows from the fact that $\boldsymbol{x_1} \cdot \boldsymbol{x_2} = \boldsymbol{x_1}^T \boldsymbol{x_2}$. □

Example 2. Consider

$$\boldsymbol{A} = \begin{bmatrix} 0 & -1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$$

We know that $\boldsymbol{A}$ has two eigenvalues $\lambda_1 = 1$ and $\lambda_2 = -2$.

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

$$\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

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

$$x_1 = -t, x_2 = -t, x_3 = t$$

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

$$\boldsymbol{x}_3 = \begin{bmatrix} -1 \\ -1 \\ 1 \end{bmatrix}$$

Vectors $\boldsymbol{x}_1$, $\boldsymbol{x}_2$, and $\boldsymbol{x}_3$ are linearly independent. According to Lemma 6, both $\boldsymbol{x_1} \cdot \boldsymbol{x_3}$ and $\boldsymbol{x_2} \cdot \boldsymbol{x_3}$ must be 0. You can verify that this is indeed the case. □

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 7. *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 2.

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

From Example 2, we know that all eigenvectors corresponding to λ_1 have the form $\begin{bmatrix} v-u \\ u \\ v \end{bmatrix}$.

For such a vector to be orthogonal to $\boldsymbol{x_1} = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$, we need:

$$\begin{aligned} (-1)(v-u)+u &= 0 \Rightarrow \\ v &= 2u \end{aligned}$$

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

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

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 7 to have length 1. □

Now we prove an important lemma about symmetric matrices.

Lemma 8. *Let $\boldsymbol{A}$ be an $n \times n$ symmetric matrix. There exist an orthogonal matrix $\boldsymbol{Q}$ such that $\boldsymbol{A} = \boldsymbol{Q}\, diag[\lambda_1, \lambda_2, ..., \lambda_n]\, \boldsymbol{Q}^{-1}$, where $\lambda_1, \lambda_2, ..., \lambda_n$ are eigenvalues of $\boldsymbol{A}$.*

Proof. From an earlier lecture, we know that given a set of linearly independent eigenvectors $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$ corresponding to eigenvalues $\lambda_1, \lambda_2, ..., \lambda_n$ respectively, we can produce $\boldsymbol{Q}$ by placing $\boldsymbol{v}_i$ as the i-th column of $\boldsymbol{Q}$, for each $i \in [1, n]$, such that $\boldsymbol{A} = \boldsymbol{Q}\, diag[\lambda_1, \lambda_2, ..., \lambda_n]\, \boldsymbol{Q}^{-1}$. From Corollary 1, we know that we can find an orthonormal set of $\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_n$. By Lemma 3, it follows that $\boldsymbol{Q}$ is an orthogonal matrix. □

Example 4. Consider once again the matrix $\boldsymbol{A}$ 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:

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

such that $\boldsymbol{A} = \boldsymbol{Q}\, diag[1, 1, -2]\, \boldsymbol{Q}^{-1}$. $\square$