Supplement: Symmetric and Hermitian Matrices

A Bunch of Definitions

Definition: A real $n \times n$ matrix $A$ is called symmetric if $A^T = A$.

Definition: A complex $n \times n$ matrix $A$ is called Hermitian if $A^* = A$, where $A^* = \overline{A^T}$, the conjugate transpose.

Definition: A complex $n \times n$ matrix $A$ is called normal if $A^* A = AA^*$, i.e. commutes with its conjugate transpose.

It is quite a surprising result that these three kinds of matrices are always diagonalizable; and moreover, one can construct an orthonormal basis (in standard inner product) for $\mathbb{R}^n/\mathbb{C}^n$, consisting of eigenvectors of $A$. Hence the matrix $P$ that gives diagonalization $A = PDP^{-1}$ will be orthogonal/unitary, namely:

Definition: An $n \times n$ real matrix $P$ is called orthogonal if $P^TP = I_n$, i.e. $P^{-1} = P^T$.

Definition: An $n \times n$ complex matrix $P$ is called unitary if $P^*P = I_n$, i.e. $P^{-1} = P^*$.

Diagonalization using these special kinds of $P$ will have special names:

Definition: A matrix $A$ is called orthogonally diagonalizable if $A$ is similar to a diagonal matrix $D$ with an orthogonal matrix $P$, i.e. $A = PDP^T$.

A matrix $A$ is called unitarily diagonalizable if $A$ is similar to a diagonal matrix $D$ with a unitary matrix $P$, i.e. $A = PDP^*$.

Then we have the following big theorems:

Theorem: Every real $n \times n$ symmetric matrix $A$ is orthogonally diagonalizable.

Theorem: Every complex $n \times n$ Hermitian matrix $A$ is unitarily diagonalizable.

Theorem: Every complex $n \times n$ normal matrix $A$ is unitarily diagonalizable.

To prove the above results, it is convenient to introduce the concept of adjoint operator, which allows us to discuss effectively the “transpose” operation in a general inner product space.

The Adjoint Operator

Let $V$ be an $n$-dimensional inner product space and let $T : V \to V$ be a linear operator. We find out that under the inner product operation, the action of $T : v \mapsto T(v)$ can be replaced/represented by another inner product action using a suitably chosen vector.

Lemma 1: Let $w \in V$ be a given vector. Then there is a unique vector $w^* \in V$ such that:

$$<T(v), w> = <v, w^*>, \quad \text{for every } v \in V. \quad (*)$$

Proof: Let $\{u_1, \ldots, u_n\}$ be an orthonormal basis for $V$. The following $w^*$ is what we want:

$$w^* = <T(u_1), w> u_1 + \ldots + <T(u_p), w> u_p = \sum_{i=1}^n <T(u_i), w> u_i.$$
Now, for \( j = 1, \ldots, n \), we check (*) for basis vector \( u_j \) first:

\[
< u_j, w^* > = < u_j, \sum_{i=1}^{n} < T(u_i), w > u_i > = \sum_{i=1}^{n} < T(u_i), w > < u_j, u_i > \\
= < T(u_j), w > < u_j, u_j > = < T(u_j), w > .
\]

So, for a general \( v \in V \), by expressing \( v = c_1 u_1 + \ldots + c_n u_n = \sum_{j=1}^{n} c_j u_j \), we have:

\[
< T(v), w > = < \sum_{j=1}^{n} c_j T(u_j), w > = \sum_{j=1}^{n} c_j < T(u_j), w > \\
= \sum_{j=1}^{n} c_j < u_j, w^* > = \sum_{j=1}^{n} c_j u_j, w^* = < v, w^* > .
\]

For the uniqueness of \( w^* \), let \( w' \in V \) be another vector with the same property, namely:

\[
< T(v), w > = < v, w^* > = < v, w' >, \quad \text{for every } v \in V.
\]

Then we take difference:

\[
< v, w^* - w' > = 0, \quad \text{for every } v \in V.
\]

In particular, this equality should be valid for \( v = w^* - w' \in V \). Thus we have:

\[
< w^* - w', w^* - w' > = 0 \quad \Rightarrow \quad || w^* - w' || = 0 \quad \Rightarrow \quad w^* = w' \quad \square
\]

**Definition:** Let \( T : V \to V \) be a linear operator. For each \( w \in V \), we define \( T^*(w) := w^* \), where \( w^* \) is the unique vector obtained in Lemma 1. This \( T^* \) is called the **adjoint** of \( T \).

**Lemma 2:** The adjoint operator \( T^* : V \to V \) is linear.

**Proof:** Straightforward checking. Let \( w_1, w_2 \in V \) and \( c, d \in \mathbb{C} \). Then for every \( v \in V \), first by definition of \( T^* \) we have:

\[
< T(v), (cw_1 + dw_2) > = < v, T^*(cw_1 + dw_2) > .
\]

But on the other hand:

\[
< T(v), (cw_1 + dw_2) > = \bar{c} < T(v), w_1 > + \bar{d} < T(v), w_2 > \\
= \bar{c} < v, T^*(w_1) > + \bar{d} < v, T^*(w_2) > \\
= < v, c T^*(w_1) + d T^*(w_2) >
\]

The above two equalities are valid for every \( v \in V \). So by the same uniqueness proof as in Lemma 1, we obtain:

\[
T^*(cw_1 + dw_2) = c T^*(w_1) + d T^*(w_2),
\]

and thus \( T^* \) is linear. \( \square \)
**Theorem 1:** Let $T, U$ be linear operators on $V$ and $k \in \mathbb{C}$. Then:

(i) $(T + U)^* = T^* + U^*$;

(ii) $(kT)^* = \bar{k}T^*$;

(iii) $(U \circ T)^* = T^* \circ U^*$;

(iv) $(T^*)^* = T$.

**Proof:** Directly from definitions. For example, the checking for (iv):

Let $v \in V$ be any vector. Then by definition:

$$<(T^*)^*(v), u> = <v, T^* (u)> = <T(v), u>,$$

for every $u \in V$.

Hence $(T^*)^*(v) = T(v)$ for any $v \in V$ and thus $(T^*)^* = T$. □

This adjoint operator $T^*$, when using matrix representation with an orthonormal basis $B$, has a simple relationship with the original linear operator $T$.

**Theorem 2:** Let $B = \{u_1, \ldots, u_p\}$ be an orthonormal basis of $V$, and let $T$ be a linear operator in $V$. Then the matrix representations of $T$ and $T^*$ relative to the orthonormal basis $B$ are given by:

$$[T]_B = \begin{bmatrix} <T(u_j), u_i> \end{bmatrix} \quad \text{and} \quad [T^*]_B = [T]^*_B.$$

**Remark:** $B$ must be orthonormal!

**Proof:** First we consider the $j$-th column of $[T]_B$, i.e. $[T(u_j)]_B$. Its entries are the $B$-coordinates of $T(u_j)$, which are exactly the coefficients in the linear combination:

$$T(u_j) = a_{1j}u_1 + \ldots + a_{nj}u_j.$$

Since $B$ is orthonormal, the $i$-th coefficient in the above linear combination can be computed effectively as:

$$<T(u_j), u_i> = a_{1j} <u_i, u_i> + \ldots + a_{nj} <u_n, u_i> = a_{ij}.$$

Thus the $(i, j)$-th entry of $[T]_B$ is given by $a_{ij} = <T(u_j), u_i>$.

Similarly the $(i, j)$-th entry of $[T^*]_B$ is given by $<T^*(u_j), u_i>$. Using the definition of adjoint operator, we have:

$$<T^*(u_j), u_i> = \overline{<u_i, T(u_j)>>} = \overline{<T(u_i), u_j>} = \bar{a}_{ji}.$$

So $[T^*]_B = [T]_B^*$ □

**Definition:** A linear operator $T : V \to V$ is called **self-adjoint** if $T^* = T$.

Thus, by Theorem 2, matrix transformation given by a symmetric/Hermitian matrix will be a self-adjoint operator on $\mathbb{R}^n/\mathbb{C}^n$, using the standard inner product.

Next we need to setup some technical lemmas for the proof of the main theorem.
Lemma 3: Let $T$ be a self-adjoint operator on $V$. Then every eigenvalue of $T$ must be real.

Proof: Let $v \neq 0$ be an eigenvector of $T$ corresponding to eigenvalue $\lambda$. We consider:

$$<T(v),v> = <\lambda v, v> = \lambda <v, v>.$$  

On the other hand, since $T^* = T$, we also have:

$$<T(v),v> = <v, T^*(v)> = <v, T(v)> = <v, \lambda v> = \bar{\lambda} <v, v>.$$  

As $<v, v> \neq 0$, we must have $\lambda = \bar{\lambda}$, i.e. $\lambda$ is real. \qed

Lemma 4: Every self-adjoint operator on $V$ has an eigenvector.

Proof: Take an orthonormal basis $B$ of $V$. Then we get a symmetric/Hermitian matrix $A = [T]_B$. By the fundamental theorem of algebra, $A$ must have an eigenvalue $\lambda \in \mathbb{C}$, and hence a corresponding eigenvector $x \in \mathbb{C}^n$. In complex case we just send this $x \in \mathbb{C}^n$ back to $v \in V$ by inverse $B$-coordinate mapping, then we will get $T(v) = \lambda v$. In real case, we apply Lemma 3 to know that this $\lambda$ must be real. Hence $x \in \mathbb{R}^n$ and we can send it back to $v \in V$ to get $T(v) = \lambda v$ again. \qed

Lemma 5: Let $W$ be a subspace of $V$ such that $T(W) \subseteq W$, i.e. $T(w) \in W$ for every $w \in W$. Then $T^*(W^\perp) \subseteq W^\perp$.

Proof: Let $z \in W^\perp$. Then for $w \in W$:

$$<w, T^*(z)> = <T(w), z> = 0 \quad \text{as} \quad T(w) \in W \text{ and } z \in W^\perp.$$  

Since the above is valid for every $w \in W$, we should have $T^*(z) \in W^\perp$. \qed

Lemma 6: Let $W$ be a subspace of an $n$-dimensional inner product space $V$. Then:

$$\dim W + \dim W^\perp = n = \dim V.$$  

Proof: Let $\{w_1, \ldots, w_k\}$ and $\{z_1, \ldots, z_{\ell}\}$ be orthogonal bases of $W$ and $W^\perp$ respectively. The lemma is proved if we can show that $S = \{w_1, \ldots, w_k, z_1, \ldots, z_{\ell}\}$ forms a basis for $V$.

Spanning $V$: For every $v \in V$, we have the orthogonal decomposition of $v$ w.r.t. $W$:

$$v = \text{proj}_W v + (v - \text{proj}_W v), \quad \text{where} \quad \text{proj}_W v \in W \text{ and } (v - \text{proj}_W v) \in W^\perp.$$  

Use the bases of $W$ and $W^\perp$ to express $\text{proj}_W v = \sum_{i=1}^k c_i w_i$ and $(v - \text{proj}_W v) = \sum_{j=1}^\ell d_j z_j$. Hence $v$ can be expressed as a linear combination of vectors in $S$.

Linearly independent: Consider the vector equation:

$$c_1 w_1 + \ldots + c_k w_k + d_1 z_1 + \ldots + d_\ell z_\ell = 0.$$  

Take inner product with $w_i$. As $\{w_1, \ldots, w_k\}$ is an orthogonal set, we have $<w_i, w_1> = 0$ for $i \neq 1$. On the other hand, since $w_1 \in W$ and all $z_j \in W^\perp$, we get $<z_j, w_1> = 0$ for all $1 \leq j \leq \ell$. So the above vector equation will become:

$$c_1 ||w_1||^2 + 0 + \ldots + 0 = <0, w_1> = 0.$$  

As $w_1 \neq 0$, we get $c_1 = 0$. Similarly for other $c_i$ and $d_j$ and they are all zeros. Thus $S$ is also linearly independent. \qed
Now we are ready to prove the main theorem.

**Diagonalizability of Symmetric and Hermitian Matrices**

**Main Theorem:** Let \( T^* = T \) be a self-adjoint linear operator on \( V \). Then \( V \) has an orthonormal basis consisting of eigenvectors of \( T \).

**Proof:** We use induction on \( n = \dim V \).

\( n = 1 \): Any non-zero vector \( v_1 \) will be an eigenvector of \( T \) since \( V = \text{Span} \{ v_1 \} \). After normalization, \( u_1 = \frac{v_1}{||v_1||} \), we obtain an orthonormal basis \( \{ u_1 \} \) of \( V \) consisting of eigenvector of \( T \).

Now, assume the statement is true for \( \dim V = k \). Next consider \( \dim V = k + 1 \).

By Lemma 4, \( T \) has an eigenvector \( u_1 \) (may assume \( ||u_1|| = 1 \)) corresponding to eigenvalue \( \lambda_1 \). Let \( W = \text{Span} \{ u_1 \} \). Note that \( T(W) = W \).

By Lemma 5, we have \( T^*(W^\perp) \subseteq W^\perp \). Since \( T^* = T \), this gives \( T(W^\perp) \subseteq W^\perp \). In other words, we can regard \( T \) as a linear operator defined on \( W^\perp \). Note that Lemma 6 says that \( \dim W^\perp = \dim V - \dim W = k \), so by induction hypothesis, there is an orthonormal basis of \( W^\perp \) consisting of eigenvectors of \( T \), say \( \{ u_2, \ldots, u_{k+1} \} \).

Since \( u_1 \in W \), \( ||u_1|| = 1 \), and \( \{ u_2, \ldots, u_{k+1} \} \) \( \subseteq W^\perp \), the combined set \( \{ u_1, u_2, \ldots, u_{k+1} \} \) is again orthonormal. This will be an orthonormal basis of \( V \) consisting of eigenvectors of \( T \).

In the case of symmetric (or Hermitian) matrix transformation, by using such an orthonormal basis of eigenvectors to construct the matrix \( P \), we will have the diagonalization \( A = PDP^{-1} \) with \( P^{-1} = P^T \) (or \( P^{-1} = P^* \)).

**Remark:** To find this \( P \), we have a more efficient method than the inductive construction in the proof of main theorem.

**Lemma 7:** Let \( T^* = T \). Then eigenvectors of \( T \) corresponding to distinct eigenvalues are orthogonal to each other.

**Proof:** Let \( T(v_1) = \lambda_1 v_1 \) and \( T(v_2) = \lambda_2 v_2 \) with \( \lambda_1 \neq \lambda_2 \). Consider on the one hand:

\[ <T(v_1), v_2> = <\lambda_1 v_1, v_2> = \lambda_1 <v_1, v_2>, \]

and on the other hand:

\[ <T(v_1), v_2> = <v_1, T^*(v_2)> = <v_1, T(v_2)> = <v_1, \lambda_2 v_2> = \lambda_2 <v_1, v_2>. \]

Since \( T \) is self-adjoint, \( \lambda_2 \) must be real, so we obtain:

\[ \lambda_1 <v_1, v_2> = \lambda_2 <v_1, v_2>. \]

As \( \lambda_1 \neq \lambda_2 \), we must have \( <v_1, v_2> = 0 \). \( \square \)

**Corollary:** Let \( T^* = T \) and let \( \{ v_{i_1} \}, \ldots, \{ v_{i_p} \} \) be orthogonal sets of eigenvectors corresponding to distinct eigenvalues \( \lambda_1, \ldots, \lambda_p \) of \( T \). Then the total collection of eigenvectors \( \{ v_{ji}; 1 \leq i \leq p \} \) is again orthogonal.

**Proof:** Exercise.
With Lemma 7 and its corollary, we only need to produce orthonormal basis for each eigenspace, which can be done by a Gram-Schmidt process. Then the total collection will be automatically orthonormal. And it is guaranteed by the main theorem that $A$ must be diagonalizable.

Remark: If $v_1, v_2$ are eigenvectors of $A$ corresponding to distinct eigenvalues, we know that $v_1 + v_2$ can never be an eigenvector of $A$. So Gram-Schmidt process should not be applied across bases for different eigenspaces.

Example: Orthogonally diagonalize the following symmetric matrix:

$$A = \begin{bmatrix} 1 & 2 & -4 \\ 2 & -2 & -2 \\ -4 & -2 & 1 \end{bmatrix}.$$  

Solution: The characteristic equation of $A$ is:

$$\det(A - \lambda I) = -\lambda^3 + 27\lambda + 54 = -(\lambda + 3)^2(\lambda - 6) = 0.$$  

So the eigenvalues are $-3, -3, 6$.

For the eigenvalue $\lambda = -3$, we solve for $\text{Nul}(A + 3I)$:

$$A + 3I = \begin{bmatrix} 4 & 2 & -4 \\ 2 & 1 & -2 \\ -4 & -2 & 4 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & \frac{1}{2} & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$  

So $\text{Nul}(A + 3I)$ has a basis $\{[1 \ 0 \ 1]^T, [-\frac{1}{2} \ 1 \ 0]^T\}$. By Gram-Schmidt process, we obtain an orthonormal basis for $\text{Nul}(A + 3I)$:

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} \\ 0 \\ \frac{1}{\sqrt{2}} \end{bmatrix}, \begin{bmatrix} -\frac{1}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \\ \frac{1}{\sqrt{18}} \end{bmatrix} \right\}.$$  

For the eigenvalue $\lambda = 6$, we solve for $\text{Nul}(A - 6I)$:

$$A - 6I = \begin{bmatrix} -5 & 2 & -4 \\ 2 & -8 & -2 \\ -4 & -2 & -5 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 0 & 1 \\ 0 & 1 & \frac{1}{2} \\ 0 & 0 & 0 \end{bmatrix}.$$  

So $\text{Nul}(A - 6I)$ has a basis $\{[-1 \ -\frac{1}{2} \ 1]^T\}$ and we obtain an orthonormal basis for $\text{Nul}(A - 6I)$:

$$\left\{ \begin{bmatrix} -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{3}} \end{bmatrix} \right\}.$$  

We construct the orthogonal matrix $P$ and diagonal matrix $D$ as:

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{18}} & -\frac{2}{3} \\ 0 & \frac{4}{\sqrt{18}} & -\frac{1}{3} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{18}} & \frac{2}{3} \end{bmatrix}, \quad D = \begin{bmatrix} -3 & 0 & 0 \\ 0 & -3 & 0 \\ 0 & 0 & 6 \end{bmatrix}.$$  

Then one can check that $A = PDP^T$.  

6
Note: The diagonalization $A = PDP^T$ is not unique, as one can have different choices of orthonormal bases for those eigenspaces with dimension greater than one. For example, the above $A$ also allows an orthogonal diagonalization $A = QDQ^T$ with:

$$Q = \begin{bmatrix} \frac{1}{3} & \frac{2}{3} & \frac{2}{3} \\ \frac{2}{3} & \frac{1}{3} & -\frac{2}{3} \\ \frac{2}{3} & \frac{2}{3} & \frac{1}{3} \end{bmatrix}.$$ 

Diagonalization of Complex Normal Matrices

Definition: A linear operator $T$ on $V$ is called normal if $T \circ T^* = T^* \circ T$.

To make the proof of main theorem also work for normal operator, we need the following technical lemma.

Lemma 8: Let $T$ be a normal operator on $V$. Then:

(i) $v$ is an eigenvector of $T$ corresponding to eigenvalue $\lambda$ 
    $\Leftrightarrow$ $v$ is an eigenvector of $T^*$ corresponding to eigenvalue $\bar{\lambda}$.

(ii) Eigenvectors corresponding to distinct eigenvalues of $T$ are orthogonal to each other.

Proof: (i) First we claim that $\|T(v)\| = \|T^*(v)\|$.

$$\|T(v)\|^2 = \langle T(v), T(v) \rangle = \langle v, TT^*(v) \rangle = \langle v, T^*T(v) \rangle = \langle v, T^*(v) \rangle = \|T^*(v)\|^2.$$ 

Then for any scalar $\lambda$, note that the operator $U = T - \lambda I$ is also normal with $U^* = T^* - \bar{\lambda} I$, so we have:

$$\|(T - \lambda I)(v)\| = \|(T^* - \bar{\lambda} I)(v)\|. \quad (\ast)$$ 

Hence:

$v$ is an eigenvector of $T$ corresponding to eigenvalue $\lambda$

$\Leftrightarrow (T - \lambda I)(v) = 0$

$\Leftrightarrow (T^* - \bar{\lambda} I)(v) = 0 \quad (\text{by } (\ast))$

$\Leftrightarrow v$ is an eigenvector of $T^*$ corresponding to eigenvalue $\bar{\lambda}$

(ii) Now let $v_1, v_2$ be eigenvectors of $T$, corresponding to distinct eigenvalues $\lambda_1 \neq \lambda_2$ respectively. Consider on the one hand:

$$\langle T(v_1), v_2 \rangle = \langle \lambda_1 v_1, v_2 \rangle = \lambda_1 \langle v_1, v_2 \rangle;$$ 

and on the other hand:

$$\langle T(v_1), v_2 \rangle = \langle v_1, T^*(v_2) \rangle = \langle v_1, \bar{\lambda}_2 v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle.$$ 

So we again obtain:

$$\lambda_1 \langle v_1, v_2 \rangle = \lambda_2 \langle v_1, v_2 \rangle.$$

As $\lambda_1 \neq \lambda_2$, we must have $\langle v_1, v_2 \rangle = 0$. □
Now, we give the proof of main theorem for normal operators.

**Main Theorem**: Let $T$ be a normal operator on a complex inner product space $V$. Then $V$ has an orthonormal basis consisting of eigenvectors of $T$.

**Proof**: We use induction on $n = \dim V$.

$n = 1$: Same as before.

Now, assume the statement is true for $\dim V = k$. Next consider $\dim V = k + 1$.

Since $V$ is a complex inner product space, $T$ will have an eigenvector $u_1$ (may assume $\|u_1\| = 1$) corresponding to eigenvalue $\lambda_1$. (*For real inner product space we might get stuck at this point.*)

By Lemma 8(i), $u_1$ is also an eigenvector of $T^*$. So if we set $W = \text{Span} \{u_1\}$, we have $T^*(W) \subseteq W$.

By Lemma 5, we have $(T^*)(W^\perp) \subseteq W^\perp$. As $(T^*)^* = T$, this means $T(W^\perp) \subseteq W^\perp$. Then we can continue the inductive argument as in the previous proof of Main Theorem.

$\square$