Inner Product Spaces: Part 2

Let $V$ be an inner product space. We let $K = \mathbb{R}$ or $\mathbb{C}$ according as $V$ is real or complex. A sequence of vectors $u_1, u_2, u_3, ...$ is said to be **orthogonal** if $< u_i, u_j >= 0$ for $i \neq j$. If, in addition, we have $||u_i|| = 1$ for all $i$ then the sequence is said to be **orthonormal**.

If $u_1, u_2, u_3, ..., u_n$ is an orthogonal sequence of non-zero vectors of $V$ then $u_1, u_2, ..., u_n$ is linearly independent. Indeed, if $a_1u_1 + a_2u_2 + \cdots + a_nu_n = 0$, we have

$$0 =< a_1u_1 + a_2u_2 + \cdots + a_nu_n, u_i >= a_i < u_i, u_i >$$

which implies that $a_i = 0$ since $< u_i, u_, >= ||u_i||^2 \neq 0$. If $u_1, u_2, ..., u_n$ is a basis of $V$ then

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n \implies < v, u_i >= a_i < u_i, u_i >$$

so that $a_i = \frac{<v,u_i>}{<u_i,u_i>}$. If $u_1, u_2, ..., u_n$ is an orthonormal basis of $V$ then

$$v = a_1u_1 + a_2u_2 + \cdots + a_nu_n \implies a_i =< v, u_i > .$$

Let $u_1, u_2, ..., u_n$ be an orthogonal sequence of non-zero vectors of $V$ and let $v \in V$. If

$$w = \frac{< v, u_1 >}{< u_1, u_1 >}u_1 + \frac{< v, u_2 >}{< u_2, u_2 >}u_2 + \cdots + \frac{< v, u_n >}{< u_n, u_n >}u_n$$

then $v - w$ is orthogonal to each vector $u_i$. If $u$ is any vector in the subspace $W$ of $V$ spanned by $u_1, u_2, ..., u_n$, we have $v - u = v - w + w - u$ with $v - w$ orthogonal to $w - u$ and so

$$||v - u||^2 = ||v - w||^2 + ||w - u||^2 \geq ||v - w||^2$$

with equality iff $u = w$. It follows that $w$ is the unique vector of $W$ which minimizes $||v - w||$. The vector $P_W(v) = w$ is called the **orthogonal projection** of $v$ on $W$. The mappping $P_W$ is a linear operator on $V$ with image $W$ and kernel the set $W^\perp$ of vectors of $V$ which are orthogonal to every vector of $W$. The subspace $W^\perp$ is called the **orthogonal complement** of $W$ in $V$. We have $V = W \oplus W^\perp$.

**Theorem 1.** *Every finite-dimensional inner product space has an orthonormal basis.*

*Proof.* Let $u_1, u_2, ..., u_n$ be a basis of the inner product space $V$. Define $v_1, v_2, ..., v_n$ inductively by

$$v_1 = u_1,\ v_{i+1} = u_{i+1} - (\frac{< u_{i+1}, v_1 >}{< v_1, v_1 >} + \frac{< u_{i+1}, v_2 >}{< v_2, v_2 >} + \cdots + \frac{< u_{i+1}, v_i >}{< v_i, v_i >}).$$

This process is well-defined since one proves inductively that

$$\mathrm{Span(v_1, ..., v_i) = Span(u_1, ..., u_i)}$$

so that, in particular, $v_i \neq 0$ for all $i$. By construction, $v_1, ..., v_n$ is orthogonal and so can be normalized to give an orthonormal basis of $V$. The above process is known as the **Gram-Schmidt Process**. □

**Corollary 2.** *An $n$-dimensional inner product space $V$ is isomorphic to $K^n$ with the ususal inner product.*

*Proof.* Let $u_1, ..., u_n$ be an orthonormal basis of $V$ and let $x, y$ be the coordinate vectors of $u, v \in V$ with respect to this basis. Then

$$\begin{aligned} < u, v > &=< x_1u_1 + x_2u_2 + \cdots + x_nu_n, y_1u_1 + y_2u_2 + \cdots + xy_nu_n \\ &= \sum_{i,j=1}^{n} x_iy_j < u_i, u_j > \\ &= x_1\overline{y}_1 + x_2\overline{y}_2 + \cdots + x_n\overline{y}_n. \end{aligned}$$

□

**Example 1.** *If* $W = \text{Span}(1, 1, 1, 1, 1)$ *in* $\mathbb{R}^4$ *then, to find an orthonormal basis for* $W^\perp$*, we first find a basis of* $W^\perp$*. The following vectors*

$$u_1 = (1, -1, 0, 0),\ \ u_2 = (1, 0, -1, 0),\ \ u_3 = (1, 0, 0, -1)$$

*are such a basis. Then*

$$\begin{aligned} v_1 &= u_1 = (1, -1, 0, 0), \\ v_2 &= u_2 - \frac{< u_2, v_1 >}{< v_1, v_1 >}v_1 = (1, 0, -1, 0) - \frac{1}{2}(1, -1, 0, 0) = \frac{1}{2}(1, 1, -2, 0), \\ v_3 &= u_3 - \frac{< u_3, v_1 >}{< v_1, v_1 >}v_1 - \frac{< u_3, v_2 >}{< v_2, v_2 >}v_2 \\ &= (1, 0, 0, -1) - \frac{1}{2}(1, -1, 0, 0) - \frac{1}{3}(1/2, 1/2, -1, 0) = \frac{1}{3}(1, 1, 1, -3). \end{aligned}$$

*form an orthogonal basis of* $W^\perp$*. Normalizing these vectors, we get*

$$(1/\sqrt{2})(1, -1, 0, 0),\ \ (1/\sqrt{6})(1, 1, -2, 0),\ \ (1/2\sqrt{3})(1, 1, 1, -3)$$

*as an orthonormal basis for* $W^\perp$*.*

If $X, Y \in K^{n\times 1}$ are the coordinate matrices of the vectors $u, v$ with respect an orthonormal basis of the inner product space $V$, we have

$$< u, v >= X^t\overline{Y}.$$

If $f = (f_1, ..., f_n)$ is an orthonormal basis of $V$ and $P = [p_{ij}]$ is the transition matrix to another basis $g = (g_1, ..., g_n)$, the $j$-th column of $P$ is $P_j = [g_j]_f$. It follows that the $ij$-th entry of $P^t\overline{P}$ is equal to

$$P_i^t\overline{P}_j =< g_i, g_j > .$$

Hence, $g$ is an orthonormal basis iff $P^t\overline{P} = I$ or, equivalently,

$$P^{-1} = \overline{P}^t.$$

Such a matrix $P$ is called an **orthogonal matrix** if $K = \mathbb{R}$ and a **unitary matrix** if $K = \mathbb{C}$.

The **adjoint** of a complex matrix $A$ is the matrix $A^* = \overline{A}^t$. We have

$$(aA + bB)^* = \overline{a}A^* + \overline{b}B^*, \quad (AB)^* = B^*A^*.$$

A complex matrix is said to be **normal** if it commutes with its adjoint, i.e., $AA^* = A^*A$. Such a matrix is necessarily square. If $A = A^*$ then $A$ is said to be **self-adjoint** or **Hermitian**.

**Theorem 3.** *The eigenvalues of a Hermitian matrix are real.*

*Proof.* If $AX = \lambda X$ with $X \neq 0$ then

$$\lambda < X,X >=< AX,X >= (AX)^t\overline{X} = X^tA^t\overline{X} = X^t\overline{AX} = X^t\overline{AX} =< X,AX >= \overline{\lambda} < X,X >$$

which implies that $\overline{\lambda} = \lambda$. □

**Theorem 4.** *Let $A$ be a normal complex matrix. Then there is a unitary matrix $U$ such that $U^{-1}AU$ is a diagonal matrix.*

*Proof.* Let $P$ be an invertible matrix such that $P^{-1}AP$ is upper triangular. Let $U$ be the unitary matrix obtained by applying the Gram-Schmidt Process to the columns of $P$ and normalizing the resulting column matrices. Then, if $P_j$ and $U_j$ are the $j$-th columns of $P$ and $U$, we have for $1 \leq j \leq n$

$$\mathrm{Span(P_1,...,P_j) = Span(U_1,...,U_n)}$$

and so

$$\begin{aligned} AU_j &= A(c_{1j}P_1 + c_{2j}P_2 + \cdots + c_{jj}P_j) \\ &= c_{1j}AP_1 + c_{2j}AP_2 + \cdots + c_{jj}AP_j \\ &= b_{1j}U_1 + b_{2j}U_2 + \cdots + b_{jj}U_j \end{aligned}$$

since $AP_j \in \mathrm{Span(P_1,...,P_j) = Span(U_1,...,U_j)}$. Hence $B = U^{-1}AU$ is upper triangular, i.e., $b_{ij} = 0$ for $j < i$. Now

$$BB^* = U^*AU(U * AU)^* = U^*AUU^*A^*U = U^*AA^*U = U^*A^*AU = U^*A^*UU^*AU = B^*B$$

and so $B = [b_{ij}]$ is also normal. Comparing the $i$-th diagonal entries of $BB^*$ and $B^*B$, we get for $1 \leq i \leq n$

$$\sum_{j \geq i} |b_{ij}|^2 = \sum_{j \leq i} |b_{ji}|^2$$

which implies by induction on $i$ that $b_{ij} = 0$ for $j > i$. Hence $B$ is a diagonal matrix. □

**Corollary 5.** *Two eigenvectors of a normal matrix with distinct eigenvalues are orthogonal.*

**Corollary 6.** *A real symmetric matrix is diagonalizable.*

**Corollary 7.** *If $A$ is a real symmetric matrix, there is an orthogonal matrix $P$ with $P^{-1}AP$ a diagonal matrix.*

Let $A$ be a normal complex matrix $n \times n$ matrix and suppose that $U$ is a unitary matrix which diagonalizes $A$. Then, if $U_j$ is the $j$-column of $U$, we have $AU_j = \lambda_j U_j$. Let $R_j = U_jU_j^*$, an $n \times n$ matrix of rank 1. The matrices $R_j$ have the following properties:

$$AR_j = \lambda_j R_j, \quad R_j^* = R_j, \quad R_j^2 = R_j, \quad R_iR_j = 0 \ (i \neq j), \quad I = \sum_{j=1}^{n} R_j.$$

To prove the last identity it suffices to prove that $X = \sum_j U_jU_j^*X$ for any column matrix $X$. But this identity follows from the fact that $U_j^*X =< X,U_j >$. Multiplying the last identity on the left by $A$, we get

$$A = \lambda_1R_1 + \lambda_2R_2 + \cdots + \lambda_nR_n.$$

If $T_A$ is the linear operator on $\mathbb{C}^n$ with matrix $A$ with respect to the standard basis, the matrix $R_i$ is the matrix (with respect to the standard basis) of the orthogonal projection of $\mathbb{C}^n$ on the one-dimensional subspace spanned by the eigenvector of $T_A$ having coordinate matrix $U_i$.

If $c_1, .., c_m$ are the distinct eigenvalues of $A$, let $Q_i$ be the sum of the $R_j$ with $\lambda_j = c_i$. The matrices $Q_i$ with $1 \leq i \leq m$ have the following properties:

$$Q_j^* = Q_j, \quad Q_j^2 = Q_j, \quad Q_i Q_j = 0 \ (i \neq j), \quad I = \sum_{j=1}^{m} Q_j,$$

$$A = c_1 Q_1 + c_2 Q_2 + \cdots + c_m Q_m.$$

This decomposition of the normal matrix $A$ is called the **spectral resolution** of $A$ and the matrices $Q_i$ are called the **projection matrices** of the resolution. The adjective spectral comes from the fact that in functional analysis the set of eigenvalues of a matrix is called the spectrum of $T$. The matrix $Q_i$ is the matrix (with respect to the standard basis) of the orthogonal projection of $\mathbb{C}^n$ on the eigenspace of $T_A$ corresponding to the eigenvalue $c_i$.

From the spectral resolution of $A$ we get

$$A^n = c_1^n Q_1 + c_2^n Q_2 + \cdots + c_m^n Q_m$$

for any natural number $n$ and hence, for any polynomial $p(\lambda)$,

$$p(A) = p(c_1)Q_1 + p(c_2)Q_2 + \cdots + p(c_m)Q_m.$$

If $p(c_i) \neq 0$ for all $i$, then $p(A)$ is invertible with

$$p(A)^{-1} = p(c_1)^{-1}Q_1 + p(c_2)^{-1}Q_2 + \cdots + p(c_m)^{-1}Q_m.$$

If $p_i(\lambda) = \prod_{j \neq i} \frac{\lambda - c_i}{c_j - c_i}$, we have $p_i(c_i) = 1$ and $p_j(c_i) = 0$ for $j \neq i$. Hence $Q_i = p_i(A)$.

If $f$ is any complex valued function whose domain is a subset of $\mathbb{C}$ and contains the set of eigenvalues of $T$, we define

$$f(A) = f(c_1)Q_1 + f(c_2)Q_2 + \cdots + f(c_m)Q_m.$$

For example, if $f(z) = e^z = \sum_{n \geq 0} z^n/n!$ then

$$e^A = e^{c_1}E_1 + e^{c_2}E_2 + \cdots + e^{c_m}E_m.$$

A Hermitian matrix is said to be **positive** if the eigenvalues are $\geq 0$ and **positive definite** if the eigenvalues are $> 0$. If $A$ is a positive Hermitian matrix, then there is a positive Hermitian matrix $B$ with $B^2 = A$, namely

$$B = \sqrt{A} = \sqrt{c_1}Q_1 + \sqrt{c_2}Q_2 + \cdots + \sqrt{c_m}Q_m.$$

We shall see later that $B$ is in fact unique.

If $A$ is a real symmetric matrix, the projection matrices of its spectral resolution are real since the eigenvalues of $A$ are real. Indeed, if

$$A = \lambda_1 R_1 + \lambda_2 R_2 + \cdots + \lambda_n R_n$$

with $R_i = P_i P_i^t$, where $P_i$ is the $i$-th column of a real orthogonal matrix $P$ with $P^{-1}AP$ equal to a diagonal matrix with $i$-th diagonal entry equal to $\lambda_i$, the matrix $Q_i$ is the sum of the $R_j$ with $\lambda_j = c_i$. The matrix $Q_i$ is the matrix (with respect to the standard basis) of the orthogonal projection of $\mathbb{R}^n$ on the eigenspace of $T_A$ corresponding to the eigenvalue $c_i$.

**Example 2.** . *If $A$ is the real symmetric matrix* $\begin{bmatrix} 2 & 1 & 1 & 1 \\ 1 & 2 & 1 & 1 \\ 1 & 1 & 2 & 1 \\ 1 & 1 & 1 & 2 \end{bmatrix}$, *we have* $(A-1)^2 = 4(A-1)$ *and so* $(A-1)(A-5) = 0$. *The eigenspace for the eigenvalue 1 is the subspace $W^\perp$ of example 1 and so has as orthonormal basis* $f_1 = (1/\sqrt{2})(1,-1,0,0)$, $f_2 = (1/\sqrt{6})(1,1,-2,0)$, $f_3 = (1/2\sqrt{3})(1,1,1,-3)$. *The eigenspace for the eigenvalue 5 is one-dimensional with basis the unit vector* $f_4 = (1/2)(1,1,1,1)$. *The matrix*

$$P = [1_{\mathbb{R}^4}]_{f,e} = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{6} & \sqrt{3}/2 & 1/2 \\ -1/\sqrt{2} & 1/\sqrt{6} & 1/2\sqrt{3} & 1/2 \\ 0 & -2/\sqrt{6} & 1/2\sqrt{3} & 1/2 \\ 0 & 0 & -3/2\sqrt{3} & 1/2 \end{bmatrix}$$

*is an orthogonal matrix with*

$$P^{-1}AP = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 5 \end{bmatrix}.$$

*If $P_i$ is the i-th column of $P$ and $R_i = P_iP_i^t$, we have*

$$R_1 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad R_2 = \frac{1}{6}\begin{bmatrix} 1 & 1 & -2 & 0 \\ 1 & 1 & -2 & 0 \\ -2 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$R_3 = \frac{1}{12}\begin{bmatrix} 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & -3 \\ -3 & -3 & -3 & 9 \end{bmatrix}, \quad R_4 = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

*We have $A = R_1 + R_2 + R_3 + 5R_4 = Q_1 + 5Q_2$, where*

$$Q_1 = R_1 + R_2 + R_3 = \frac{1}{4}\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}, \quad Q_2 = R_4 = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

*We thus have $A^n = Q_1 + 5^nQ_2$ for all $n \in \mathbb{Z}$ and, if $B = \sqrt{A} = Q_1 + \sqrt{5}Q_2$, we have $B^2 = A$.*

As another application of this consider $e^{tA} = e^tQ_1 + e^{5t}Q_2$ with $t \in \mathbb{R}$. Then $e^{tA}$ is a matrix $C = [c_{ij}(t)]$ whose entries are differentiable real valued functions $c_{ij}(t)$ of $t$. If we define the derivative of $C = C(t)$ to be $\frac{dC}{dt} = C'(t) = [c'_{ij}(t)]$, we have $\frac{d}{dt}e^{tA} = e^tQ_1 + 5e^{5t}Q_2 = Ae^{tA}$. This can be used to solve the system of differential equations

$$\begin{aligned} \frac{dx_1}{dt} &= 2x_1 + x_2 + x_3 + x_4 \\ \frac{dx_2}{dt} &= x_1 + 2x_2 + x_3 + x_4 \\ \frac{dx_3}{dt} &= x_1 + x_2 + 2x_3 + x_4 \\ \frac{dx_4}{dt} &= x_1 + x_2 + x_3 + 2x_4 \end{aligned}$$

Indeed, writing this system in the form $\frac{dX}{dt} = AX$, it is an easy exercise to prove that

$$X = e^{tA} \begin{bmatrix} a \\ b \\ c \\ d \end{bmatrix}$$

is the unique solution with $X(0) = [a, b, c, d]^t$.