The Spectral Theorem

Let V be a finite-dimensional inner-product space. The mapping $b_R : V \to V^*$ defined by $b_R(v) = \phi_v$, where $\phi_v(u) =< u, v >$, is bijective. It is linear if V is real and conjugate linear if V is complex. If $T : U \to V$ is a linear mapping of finite-dimensional inner-product spaces, then there is a unique mapping $T^* : V \to U$ such that

$$T^t(\pi_v) = \phi_{T*(v)}.$$

Since $T^t(\phi_v) = \phi_v \circ T$, this is equivalent to

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

for all $u \in U, v \in V$. The mapping T^* is called the **adjoint** of T. Since

$$\begin{aligned} < u, T^*(c_1 v_v + c_2 v_2) > &=< T(u), c_1 v_v + c_2 v_2) > \\ &= \bar{c}_1 < T(u), v_1) + \bar{c}_1 2 < T(u), v_2) \\ &= \bar{c}_1 < u, T^*(v_1)) + \bar{c}_1 2 < u, T^*(v_2)) \\ &=< u, c_1 T^*(v_1) + c_2 T^*(v_2), \end{aligned}$$

we see that T^* is a linear mapping. We also have

$$(aS + bT)^* = \bar{a}S^* + \bar{b}T^*, \quad T^{**} = T.$$

The proof of this is left as an exercise.

If e, f are orthonormal bases of U, V respectively, we have

$$< T(u), v >= [T(u)]_f^t \overline{[v]}_f = X^t A^t \overline{Y} = X^t \overline{A^* Y}$$

where $X = [u]_e, Y = [v]_f, A = [T], A = [T]_e^f, A^* = \overline{A}^t$. Since $< u, T^*(v) >= X^t \overline{BY}$, where $B = [T^*]_f^e$ we see that $B = A^*$. Thus, for orthonormal bases e, f, we have

$$[T^*]_f^e = ([T]_e^f)^*.$$

The matrix A^* is called the adjoint of A. If T is a linear operator on V then T is said to be **self-adjoint** if $T = T^*$ and **normal** if $T^*T = TT^*$.

For example, if $A \in \mathbb{C}^{n \times n}$, then the linear operator T_A on $\mathbb{C}^{n \times 1}$ defined by $T_A(X) = AX$ is self-adjoint if and only if A is Hermitian. If $A \in \mathbb{R}^{n \times n}$, then the linear operator T_A on $\mathbb{R}^{n \times 1}$ defined by $T_A(X) = AX$ is self-adjoint if and only if A is symmetric.

If $T : U \to V$ is a linear mapping of inner product spaces then if a linear mapping $S : V \to U$ satisfies

$$< T(u), v >=< u, S(v) >$$

for all $u \in U, v \in V$ it is unique as is called the adjoint of T and denoted by T^*. For example, in the space ℓ^2, the adjoint of the left-shift operator L is the right-shift operator R since

$$< L(x), y >= \sum_{n=0}^{\infty} x_{n+1} \overline{y}_n = \sum_{n=1}^{\infty} x_n \overline{y}_{n-1} =< x, R(y) > .$$

A linear operator T on an inner-product space V is said to be self-adjoint if $T = T^*$ which is equivalent to

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

for all $u, v \in V$.

A linear operator on an inner-product space V is said to be **orthogonally diagonalizable** if there is a basis of V consisting of eigenvectors of T.

Theorem 1 (Spectral Theorem). *Let T be a linear operator on a finite-dimensional inner product space V. Then T is orthogonally diagonalizable if and only if T is normal.*

Proof. ($\Longrightarrow$) If $e = (e_1, \ldots, e_n)$ is an orthonormal basis of eigenvectors of T then

$$[T]_e = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}, \quad [T^*]_e = \begin{bmatrix} \overline{\lambda}_1 & 0 & \cdots & 0 \\ 0 & \overline{\lambda}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \overline{\lambda}_n \end{bmatrix}$$

which shows that T, T^* commute.

($\Longleftarrow$) We proceed by induction on $n = \dim(V)$, the case $n = 0$ being trivial. Let λ be an eigenvalue of T and let $W = \mathrm{Ker(T - \lambda)}$. Then W is T-invariant which implies that W is T^*-invariant since T and T^* commute. If $w \in W$, $v \in W^\perp$ then

$$< T(v), w >=< v, T^*(w) >= 0$$

which implies that $W^\perp$ is T-invariant. If S is the restriction T to $W^\perp$ then $S*$ is the restriction of $T*$ to $S^\perp$ so that S is a normal operator on the inner-product space $W^\perp$. Since $\dim(W^\perp) < n$, our inductive hypothesis implies that $W^\perp$ has an orthogonal basis B_1 consisting of eigenvectors of S. Since eigenvectors of S are also eigenvectors of T and $V = W \oplus W^\perp$ we obtain an orthogonal basis B of V consisting of eigenvectors of T eigenvectors of T by taking $B = B_1 \cup B_2$, where B_2 is an orthogonal basis of W. $\square$

Corollary 2. *If T is a normal operator then eigenvectors of T corresponding to distinct eigenvalues are orthogonal.*

This is true even if V is infinite-dimensional. Indeed, if T is normal we have

$$||T(u)||^2 =< T(u), T(u) >=< u, T^*T(u) >=< u, TT^*(u) >=< T^*(u), T^*(u) >= ||T^*(u)||^2$$

Thus $||(T - \lambda)(u)|| = ||(T^* - \overline{\lambda})(u)||$ so that $T(u) = \lambda u \iff T^*(u) = \overline{\lambda} u$. Hence if $T(u) = \lambda u$ and $T(v) = \mu v$ we have

$$\lambda < u, v >=< \lambda u, v >=< T(u), v >=< u, T^*(v) >=< u, \overline{\mu} v = \mu < u, v >$$

which implies $(\lambda - \mu) < u, v >= 0$ and hence $< u, v >= 0$.

Corollary 3. *If $A \in \mathbb{C}^{n \times n}$ is a normal matrix ($AA^* = A^*A$) then there is an invertible matrix $P \in \mathbb{C}^{n \times n}$ whose columns are orthonormal for the standard inner product on $\mathbb{C}^{n \times n}$ such that $P^{-1}AP$ is a diagonal matrix.*

The columns of $P \in \mathbb{C}^{n\times n}$ are orthonormal if and only if $P^T\overline{P} = I$ which is equivalent to $P^{-1} = P^*$. Such a complex matrix is called **unitary**. Two complex matrices A, B are said to be unitarily equivalent if there exists a unitary matrix P such that $B = P^{-1}AP$ or, equivalently, $B = P^*AP$.

Corollary 4. *If T is self-adjoint then the eigenvalues of T are real.*

This is true even if V is infinite-dimensional. Indeed, if $T(u) = \lambda u$ and $u \neq 0$ then

$$\lambda < u, u >=< \lambda u, u >=< T(u), u >=< u, T(u) >=< u, \lambda u >= \overline{\lambda} < u, u >$$

which implies $\lambda = \overline{\lambda}$ on cancelling $< u, u >= ||u||^2$ which is $\neq 0$.

Corollary 5. *If T is a linear operator on a finite dimensional real inner product space then T is orthogonally diagonalizable if and only if T is self-adjoint.*

Corollary 6. *If A is a real symmetric $n \times n$ matrix there exists an invertible real matrix P with orthonormal columns with respect to the standard inner product on $\mathbb{R}^{n\times n}$ such that $P^{-1}AP$ is a diagonal matrix.*

The columns of $P \in \mathbb{R}^{n\times n}$ are orthonormal if and only if $P^TP = I$ which is equivalent to $P^{-1} = P^t$. Such a real matrix is called **orthogonal**. Two real matrices A, B are said to be orthogonally equivalent if there exists an orthogonal matrix P such that $B = P^{-1}AP$ or, equivalently, $B = P^tAP$.

If T is a diagonalizable operator on a finite-dimensional vector space V and $\mu_1, \ldots, \mu_s$ are the distinct eigenvalues of T then

$$V = \mathrm{Ker(T - \mu_1)} \oplus \mathrm{Ker(T - \mu_2)} \oplus \cdots \oplus \mathrm{Ker(T - \mu_s)}.$$

Hence any $v \in V$ can be uniquely written in the form $v = v_1 + v_2 + \cdots + v_s$ with $v_i \in V_i = \mathrm{Ker(T - \mu_i)}$. The operators Q_i on V defined by $Q_i(v) = v_i$ are linear and satisfy

$$1 = Q_1 + Q_2 + \cdots + Q_n, \quad Q_iQ_j = \delta_{ij}.$$

Since $TQ_i = \mu_iQ_i$, we have

$$T = \mu_1Q_i + \mu_2Q_2 + \cdots + \mu_sQ_s.$$

This is the **spectral decomposition** of T. If V is an inner product space and V is self-adjoint then the operator Q_i is orthogonal projection onto V_i. This operator is self-adjoint.

If $p(X)$ is any polynomial in X then

$$p(T) = p(\mu_1)Q_1 + p(\mu_2)Q_2 + \cdots + p(\mu_s)Q_s.$$

If $c_i^m = \mu_i$ for $1 \leq i \leq s$ and

$$S = c_1Q_1 + c_2Q_2 + \cdots + c_sQ_s$$

then $S^m = T$.

If A is a real or complex $n \times n$ matrix which is orthogonally diagonalizable and $u_1, \ldots, \mu_n$ is an orthogonal basis of eigenvectors with $Au_i = \lambda_iu_i$ then $P_i = u_iu_i^*$ is the matrix of orthogonal projection onto $\mathrm{Span(u_i)}$ so that

$$I = P_1 + P_2 + \cdots + P_n, \quad P_iP_j = \delta_{ij}I, \quad AP_i = AP_i.$$

Hence $A = \lambda_1 P_1 + \lambda_2 P_2 + \cdots + \lambda_n P_n$ is a spectral decomposition of A into a linear combination of rank 1 self-adjoint matrices P_i. To get the spectral decomposition of A let $\mu_1, \ldots, \mu_s$ be the distinct eigenvalues of A and let

$$Q_i = \sum_{\lambda_j = \mu_i} P_j.$$

Then $A = \mu_1 Q_1 + \mu_2 Q_2 + \cdots + \mu_s Q_s$ is the spectral decomposition of A.

Example 1. *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 has as orthonormal basis $f_1 = (1/\sqrt{2})(1,-1,0,0)^t$, $f_2 = (1/\sqrt{6})(1,1,-2,0)^t$, $f_3 = (1/2\sqrt{3})(1,1,1,-3)^t$. The eigenspace for the eigenvalue 5 is one-dimensional with basis the unit vector $f_4 = (1/2)(1,1,1,1)^t$. 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 $P_i = f_i f_i^t$, we have

$$P_1 = \frac{1}{2}\begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \quad P_2 = \frac{1}{6}\begin{bmatrix} 1 & 1 & -2 & 0 \\ 1 & 1 & -2 & 0 \\ -2 & -2 & 4 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

$$P_3 = \frac{1}{12}\begin{bmatrix} 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & -3 \\ 1 & 1 & 1 & -3 \\ -3 & -3 & -3 & 9 \end{bmatrix}, \quad P_4 = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

We have $A = P_1 + P_2 + P_3 + 5P_4 = Q_1 + 5Q_2$, where

$$Q_1 = P_1 + P_2 + P_3 = \frac{1}{4}\begin{bmatrix} 3 & -1 & -1 & -1 \\ -1 & 3 & -1 & -1 \\ -1 & -1 & 3 & -1 \\ -1 & -1 & -1 & 3 \end{bmatrix}, \quad Q_2 = P_4 = \frac{1}{4}\begin{bmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{bmatrix}.$$

We thus have $A^n = Q_1 + 5^n Q_2$ 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 $e^{tA} = e^t Q_1 + e^{5t} Q_2$ 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 $\frac{d}{dt} e^{tA} = e^t Q_1 + 5e^{5t} Q_2 = Ae^{tA}$. 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 $X(0) = [a, b, c, d]^t$.

The exponential matrix e^A can be defined for any real or complex square matrix A by

$$e^A = I + A + \frac{A^2}{2} + \cdots + \frac{A^n}{n!} + \cdots .$$

One can show that the series converges for the usual norm on $\mathbb{R}^{n \times n}$.

We now give an application of the Spectral Theorem to real quadratic forms. If q is a quadratic form on a real inner-product space V of dimension n then $q(u) =< T(u), u >$ for a unique self-adjoint operator T on V. Indeed, if e is an orthonormal basis of V and $X = [u]_e$, we have $q(u) = X^t AX$ for a unique symmetric matrix A. We define T to be the linear operator on V with matrix $[T]_e = A$. This definition of T is independent of the choice of orthonormal basis. Indeed, if f is another orthonormal basis and $Y = [u]_f$ then $X = PY$ with P the orthogonal matrix whose columns are $[f_1]_e, [f_2]_e, \ldots, [f_n]_e$ then $q(u) = Y^t P^t APY = Y^t BY$ with $B = P^t AP = P^{-1}AP = [T]_f$. By the Spectral Theorem, we can choose f so that $[T]_f$ is a diagonal matrix. In this case

$$q(u) = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2$$

where $Y = [y_1, y_2, \ldots, y_n]$ and $T(f_i) = \lambda_t f_i$. If m is the smallest eigenvalue and M the largest we have

$$m||u||^2 = q(u) \leq M||u||^2$$

so that m, M are the minimum values of q on the unit sphere $||u|| = 1$. These values are attained if and only if u is an eigenvector of T. For example, the minimum and maximum values of the quadratic form $q(X) = X^t AX$, where A is the matrix in Example 1, are 1 and 5 respectively.

If $A \in \mathbb{R}^{m \times n}$ then $B = A^t A$ is a symmetric matrix. The quadratic form

$$q(X) = X^t B X = (AX)^t AX = ||AX||^2$$

is positive so that the eigenvalues λ_i of B are ≥ 0. If A is a square matrix then the λ_i are just the squares of the eigenvalues of A. If M is the largest of the eigenvalues λ_i then $||AX|| \leq \sqrt{M}||X||$ with equality if and only if X is an eigenvector of $A^t A$. If we define the norm of A to be $||A|| = \sqrt{M}$ then

$$||A|| = 0 \implies A = 0, \quad ||cA|| = |c|||A||, \quad ||A_1 + A_2|| \leq ||A_1|| + ||A_2||, \quad ||A_1 A_2|| \leq ||A_1||||A_2||.$$

Such a norm is called a **matrix norm**.

Let $f = (f_1, \ldots, f_n)$ be an orthonormal basis of eigenvectors of B. Suppose that

$$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_r > 0$$

are the non-zero eigenvalues of B. For $1 \leq i \leq r$, let $\sigma_i = \sqrt{\lambda_i}$ and let $g_i = \frac{1}{\sigma_i} A f_i$. then $A f_i = \sigma_i g_i$ and

$$< g_i, g_j >= \frac{1}{\lambda_i} < Af_i, Af_j >= \frac{1}{\lambda_i} < A^t A f_i, f_j >=< f_i, f_j >= \delta_{ij}$$

so that $g_1, \ldots, g_r$ is an orthonormal set of vectors in $\mathbb{R}^m$. If we set $E_i = g_i f_i^t$ then E_i is of rank 1 and

$$A = \sigma_1 E_1 + \sigma_2 E_2 + \cdots + \sigma_r E_r.$$

This is the **singular value decomposition** of A. The σ_i are the singular values of A. We have

$$||A - (\sigma_1 E_1 + \sigma_2 E_2 + \cdots + \sigma_s E_s)|| \leq \sigma_{s+1}$$

for $1 \leq s < r$. One can show that $\sigma_1 E_1 + \sigma_2 E_2 + \cdots + \sigma_s E_s$ is the best approximation to A by a matrix of rank s.

Problem 1. *Prove that*

$$\frac{1}{\sigma_1} E_1^t + \frac{1}{\sigma_2} E_2^t + \cdots + \frac{1}{\sigma_r} E_r^t$$

is the generalized inverse of A.