Math 350 Fall 2011
Notes about the Spectral theorem

1. Introduction

The goal of this article is to prove the spectral theorem for three classes of operators on a finite dimensional inner product space: namely the real symmetric operators, the (complex) self-adjoint operators, and the (complex) normal operators. Before we state the theorems, we recall the following definitions.

Suppose V is a (real or complex) finite dimensional inner product space. A basis of V is said to be an *orthonormal basis*[1] if and only if the norm of every vector in the basis is 1, and the inner product between any two distinct vectors from the basis is 0.

If V is a real inner product space and $T\colon V \to V$ is a linear operator, we define the *transpose* of T to be the map $T^t\colon V \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^t w\rangle \quad \text{for all } v, w \in V.$$

If V is a complex inner product space and $T\colon V \to V$ is a linear operator, we define the *adjoint* of T to be the map $T^*\colon V \to V$ such that

$$\langle Tv, w\rangle = \langle v, T^* w\rangle \quad \text{for all } v, w \in V.$$

If V is a real inner product space, and $T\colon V \to V$ is linear, then T is said to be *symmetric* if and only if $T^t = T$.

If V is a complex inner product space, and $T\colon V \to V$ is linear, then T is said to be *self-adjoint* (or *Hermitian*) if and only if $T^* = T$, and T is said to be *normal* if and only if $TT^* = T^*T$.

Note that every symmetric operator on a finite dimensional real inner product space extends to a self-adjoint operator on the complexification of the real inner product space (c.f. the first proof of (i) in Section 4), and every self-adjoint operator on a complex inner product space is normal.

There are three main theorems in this article. The first is a characterization of symmetric operators on real inner product spaces:

Theorem 1. *Suppose V is a finite dimensional real inner product space, and $T\colon V \to V$ is linear. Then the following are equivalent:*

(a) T is symmetric;
(b) V has an orthonormal basis that consists of eigenvectors of T;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal.

[1] Some people call an orthonormal basis of a complex inner product space a *unitary* basis. We will not adopt this terminology.

Similarly, we have the following characterization of self-adjoint operators on complex inner product spaces:

Theorem 2. *Suppose V is a finite dimensional complex inner product space, and $T\colon V \to V$ is linear. Then the following are equivalent:*

(a) T is self-adjoint;
(b) V has an orthonormal basis that consists of eigenvectors of T, and all eigenvalues of T are real;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal with real entries.

Finally we have the following characterization of normal operators on complex inner product spaces:

Theorem 3. *Suppose V is a finite dimensional complex inner product space, and $T\colon V \to V$ is linear. Then the following are equivalent:*

(a) T is normal;
(b) V has an orthonormal basis that consists of eigenvectors of T;
(c) V has an ordered orthonormal basis such that the matrix representation of T with respect to this basis is diagonal.

It is actually beneficial to work in this abstract setting even if one is only interested in the corresponding results about matrices. As a result, we keep this level of abstraction for the moment, and only give the matrix versions of the above theorems at the end of the article.

To put these theorems in perspective, we note the role played by orthogonality here. For instance, suppose V is a finite dimensional complex inner product space. Then Theorem 3 can be seen as a characterization of linear operators $T\colon V \to V$ for which V admits an *orthonormal* basis consisting of eigenvectors of T. The class of such linear operators turns out to be quite small, as one sees from the theorem; in fact the set of normal operators on V is nowhere dense in the set of all linear operators on V. On the other hand, if we drop the assumption of orthogonality on the basis of V, i.e. if we are interested instead in the set of linear operators on V for which V admits a basis (not necessarily orthonormal) consisting of eigenvectors of T, then we will get the set of all diagonalizable linear operators on V, and this set turns out to be dense in the set of all linear operators on V.

2. The easy implications

In each of the above theorems, (b) and (c) are almost a tautology. We leave its proof to the reader. Also, it is very easy to show that (b) implies (a). For instance, let's work in the setting of Theorem 2. Suppose V is a complex inner product space, and V has an orthonormal basis $\{v_1, \dots, v_n\}$ that consists of eigenvectors of T, say $Tv_j = \lambda_j v_j$ for all $j = 1, \dots, n$. Suppose also that all the λ_j's are real. Then for any $j, k = 1, \dots, n$, we have

$$\langle v_j, T^* v_k \rangle = \langle T v_j, v_k \rangle = \lambda_j \langle v_j, v_k \rangle = \begin{cases} \lambda_j & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} .$$

It follows that

$$\langle T^* v_k, v_j \rangle = \overline{\langle v_j, T^* v_k \rangle} = \begin{cases} \lambda_k & \text{if } j = k \\ 0 & \text{otherwise} \end{cases} ,$$

since the $\lambda_1, \dots, \lambda_n$ are real. Thus

$$T^* v_k = \sum_{j=1}^{k} \langle T^* v_k, v_j \rangle v_j = \lambda_k v_k = T v_k \quad \text{for all } k = 1, \dots, n.$$

Since $\{v_1, \dots, v_n\}$ form a basis of V, it follows that $T^* = T$.

A similar proof that (b) implies (a) is possible in Theorems 1 and 3. It remains to show that (a) implies (b) for each of the theorems above. We first do this in the setting of Theorem 2, and then give the modifications necessary to prove Theorem 1. Finally we prove that (a) implies (b) in the setting of Theorem 3.

3. Proof of Theorem 2

Suppose $T : V \to V$ is a self-adjoint linear operator on a finite dimensional complex inner product space V. First we prove the following proposition:

Proposition 1. *All the eigenvalues of T are real.*

Proof. Suppose $\lambda \in \mathbb{C}$ is an eigenvalue of T, and $v \neq 0$ is the corresponding eigenvector of T. Then $Tv = \lambda v$, so

$$\lambda \langle v, v \rangle = \langle Tv, v \rangle = \langle v, Tv \rangle = \overline{\lambda} \langle v, v \rangle.$$

Since $\|v\| \neq 0$, we get

$$\lambda = \overline{\lambda},$$

and thus λ is real. □

Next, we want to prove that V has an orthonormal basis consisting of eigenvectors of T. The proof here is by induction on the dimension of V. It comes in two steps. First, given T as above, we find an eigenvector v_1 of T. Without loss of generality we may assume that $\|v_1\| = 1$. Let W be the orthogonal complement of v_1 in V. We will then show[2] that W is a T-invariant subspace of V, and that the restriction of T to W is still self-adjoint. One can then invoke our induction hypothesis, and conclude that W has an orthonormal basis $\{v_2, \dots, v_n\}$ that consists of eigenvectors of T. Then one can check that $\{v_1, \dots, v_n\}$ is an orthonormal basis of V consisting of eigenvectors of T.

As a result, we only need to prove the following:

[2]This is a crucial place where orthogonality comes in; it is not just any complement of the span of v_1 in V that is T-invariant, but the orthogonal complement of v_1 that is T-invariant.

(i) Every self-adjoint linear operator $T\colon V \to V$ on a finite dimensional complex inner product space V has an eigenvector;

(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W \to W$ is self-adjoint.

We begin with (ii). If v_1 is an eigenvector of T, say with eigenvalue λ ($\in \mathbb{R}$ by the previous proposition), and if W is the orthogonal complement, then for any $w \in W$, we have

$$\langle Tw, v_1 \rangle = \langle w, T^* v_1 \rangle = \langle w, T v_1 \rangle = \lambda \langle w, v_1 \rangle = 0.$$

Thus Tw is in W for all $w \in W$. This proves that W is T-invariant.

Next, since $T^* = T$, W is also T^*-invariant. Thus if U is the restriction of T to W, then, U^* is the restriction of T^* to W; in fact if $w_1, w_2 \in W$, then

$$\langle w_1, U^* w_2 \rangle = \langle U w_1, w_2 \rangle = \langle T w_1, w_2 \rangle = \langle w_1, T w_2 \rangle = \langle w_1, U w_2 \rangle,$$

which shows $U^* w_2 = U w_2$ for all $w_2 \in W$. (Here one uses $U w_2 \in W$.) It follows that $U^* = U$, and thus the restriction of T to W is self-adjoint.

We next prove (i). In fact we will give three different proofs.

The easiest proof is to invoke the fundamental theorem of algebra, which says that every complex polynomial has a root in the complex plane. In particular, since V is finite dimensional, there exists some $\lambda_1 \in \mathbb{C}$ such that $\det(T - \lambda_1 I) = 0$. This says T has an eigenvalue λ_1, and thus T has an eigenvector v_1 corresponding to λ_1.

Alternatively, to prove (i), we proceed by a *variational method*: Let S be the unit sphere in V, i.e. $S = \{v \in V \colon \|v\| = 1\}$. It is a compact subset of V by the finite dimensionality of V. We define a function $F\colon S \to \mathbb{R}$ by

$$F(v) = \langle Tv, v \rangle, \quad v \in S.$$

(Note that $\langle Tv, v \rangle$ is always real since T is self-adjoint; in fact, we have $\langle Tv, v \rangle = \langle v, Tv \rangle = \overline{\langle Tv, v \rangle}$ for all $v \in V$.) The function $F\colon S \to \mathbb{R}$ is smooth; hence by compactness of S, it achieves a maximum at some $v_1 \in S$. Thus for any vector $w \in V$, the function $t \in \mathbb{R} \mapsto F\left(\frac{v_1 + tw}{\|v_1 + tw\|}\right) \in \mathbb{R}$ achieves a maximum at $t = 0$. It follows that

$$\left.\frac{d}{dt}\right|_{t=0} F\left(\frac{v_1 + tw}{\|v_1 + tw\|}\right) = 0.$$

But the left hand side of this equation is equal to

$$\left.\frac{d}{dt}\right|_{t=0} \frac{1}{\|v_1 + tw\|^2} \langle T(v_1 + tw), v_1 + tw \rangle$$

which in turn is given by

$$-\frac{\langle v_1, w \rangle + \langle w, v_1 \rangle}{\|v_1\|^4} \langle T v_1, v_1 \rangle + \frac{1}{\|v_1\|^2} (\langle T v_1, w \rangle + \langle T w, v_1 \rangle),$$

i.e.

$$-(\langle v_1, w \rangle + \langle w, v_1 \rangle) \langle T v_1, v_1 \rangle + \langle T v_1, w \rangle + \langle T w, v_1 \rangle. \tag{1}$$

Since we are working on a complex inner product space, and since $T^* = T$, this is equal to

$$-2\langle Tv_1, v_1\rangle \text{Re}\langle v_1, w\rangle + 2\text{Re}\langle Tv_1, w\rangle.$$

Writing $\lambda_1 = \langle Tv_1, v_1\rangle \in \mathbb{R}$, we then get

$$\text{Re}\,\langle Tv_1 - \lambda_1 v_1, w\rangle = 0$$

for all $w \in V$. By replacing in this identity w by iw, we get also

$$\text{Im}\,\langle Tv_1 - \lambda_1 v_1, w\rangle = 0$$

for all $w \in V$. It follows that

$$\langle Tv_1 - \lambda_1 v_1, w\rangle = 0$$

for all $w \in V$; thus

$$Tv_1 = \lambda_1 v_1,$$

and v_1 is an eigenvector of T. (Note that $\|v_1\| = 1$, so in particular $v_1 \neq 0$.) This proves (i). (Incidentally, by iterating this argument on the orthogonal complements of the span of the eigenvectors we found, this gives another proof that all eigenvalues of T are real, since they are all of the form $\langle Tv, v\rangle$ where v is an eigenvector of T, and $\langle Tv, v\rangle \in \mathbb{R}$ for all $v \in V$.)

We remark that the variational approach works here only because we are dealing with a special class of linear operators, namely the self-adjoint ones. In general, if $T\colon V \to V$ is just any linear operator (not necessarily self-adjoint) on a finite dimensional complex inner product space, it is not true that the supremum of $|\langle Tv, v\rangle|$ over all $v \in V$ with $\|v\| = 1$ is achieved at an eigenvector of T. For example, if $T\colon \mathbb{C}^2 \to \mathbb{C}^2$ is the linear map given by $T(z_1, z_2) = (z_1 + z_2, z_2)$, then $|\langle T(z), z\rangle| = ||z|^2 + \overline{z_1} z_2|$ for $z = (z_1, z_2)$, and as z varies over $\mathbb{C}^2$ with $|z| = 1$, the supremum is achieved when z is a (complex) multiple of $(1, 1)$. Nonetheless, such z is not an eigenvector of T; the only eigenvectors of T are (complex) multiplies of $(1, 0)$. (It is true, however, that if $T\colon V \to V$ is a normal operator on a finite dimensional complex inner product space, then the supremum of $|\langle Tv, v\rangle|$ over all $v \in V$ with $\|v\| = 1$ is achieved at an eigenvector of T; one can see this rather easily once one has proved Theorem 3 for normal operators.)

Finally, we give a proof of (i) by comparing the *operator norm* and the *numerical radius* of T. We need some preparation.

3.1. **Operator norm.** In this subsection, suppose V is a finite dimensional[3] complex inner product space, and $T\colon V \to V$ is a linear operator. The *operator norm* of T is defined to be

$$\|T\| = \sup_{\|v\|=1} \|Tv\|.$$

We need a few propositions about the operator norm:

Proposition 2. $\|T\| = \sup_{\|v\|=\|w\|=1} |\langle Tv, w\rangle|.$

[3]In fact all results in this subsection holds, if V is a complex inner product space (possibly infinite dimensional), and if $T\colon V \to V$ is a *continuous* linear operator on V.

Proof. Let $\sigma(T)$ be the right hand side of the above identity. First, by the Cauchy-Schwarz inequality, if $v, w \in V$ are so that $\|v\| = \|w\| = 1$, then

$$\sigma(T) = \sup_{\|v\|=\|w\|=1} |\langle Tv, w\rangle| \leq \|Tv\|\|w\| \leq \|T\|,$$

the latter inequality following from the definition of $\|T\|$. Thus $\sigma(T) \leq \|T\|$. Next, it is easy to see that

$$|\langle Tv, w\rangle| \leq \sigma(T)\|v\|\|w\|$$

for all $v, w \in V$; this is because if v, w are both not zero, then one can write $\langle Tv, w\rangle$ as $\langle T(\frac{v}{\|v\|}), \frac{w}{\|w\|}\rangle\|v\|\|w\|$. Then $\frac{v}{\|v\|}, \frac{w}{\|w\|}$ have norm 1, and by definition of $\sigma(T)$ we get the desired inequality. The inequality holds trivially if one of v or w is zero. Thus taking $w = Tv$, we get

$$\|Tv\|^2 = \langle Tv, Tv\rangle \leq \sigma(T)\|v\|\|Tv\|$$

for all $v \in V$. In other words,

$$\|Tv\| \leq \sigma(T)\|v\|$$

for all $v \in V$ (if $\|Tv\| \neq 0$ we just divide by $\|Tv\|$ from both sides; if $\|Tv\| = 0$ this inequality is trivial). Thus for any $v \in V$ with $\|v\| = 1$, we have $\|Tv\| \leq \sigma(T)$, from which it follows that $\|T\| \leq \sigma(T)$. Putting these together we get $\|T\| = \sigma(T)$. □

Proposition 3. $\|T\|^2 = \|T^*T\|$

Proof. First,

$$\|T\|^2 = \sup_{\|v\|=1} \langle Tv, Tv\rangle = \sup_{\|v\|=1} \langle T^*Tv, v\rangle \leq \sup_{\|v\|=1} \|T^*Tv\|\|v\| \leq \|T^*T\|.$$

The first equality follows from $\|T\|^2 = \sup_{\|v\|=1} \|Tv\|^2$, and the first inequality is the Cauchy-Schwarz. Next, by the previous proposition,

$$\|T^*T\| = \sup_{\|v\|=\|w\|=1} |\langle T^*Tv, w\rangle| = \sup_{\|v\|=\|w\|=1} |\langle Tv, Tw\rangle| \leq \sup_{\|v\|=\|w\|=1} \|Tv\|\|Tw\| = \|T\|^2,$$

where the first inequality is again Cauchy-Schwarz. Together we conclude that $\|T\|^2 = \|T^*T\|$. □

Proposition 4. *If T is self-adjoint, then $\|T\|^{2^k} = \|T^{2^k}\|$ for all $k \in \mathbb{N}$. In particular, $\|T\|^2 = \|T^2\|$.*

Proof. The proof is by induction on k. First, if T is self-adjoint, i.e. if $T^* = T$, then the previous proposition says

$$\|T\|^2 = \|T^2\|.$$

Next, if T is self-adjoint, then so is T^2, so by the result we have just obtained (for $k = 1$), we get

$$\|T\|^4 = (\|T\|^2)^2 = \|T^2\|^2 = \|(T^2)^2\| = \|T^4\|.$$

In general, if the statement has been proved a certain $k \in \mathbb{N}$, then for any self-adjoint T, we have

$$\|T\|^{2^{k+1}} = (\|T\|^2)^{2^k} = \|T^2\|^{2^k} = \|(T^2)^{2^k}\| = \|T^{2^{k+1}}\|,$$

where the third equality follows from the induction hypothesis (and that T^2 is self-adjoint). This completes our induction. □

3.2. **Numerical radius.** Next, we introduce the numerical radius of an operator. Again in this section, V is a finite dimensional[4] complex inner product space, and $T\colon V \to V$ is a linear operator.

The *numerical radius* of T is then defined to be

$$r(T) = \sup_{\|v\|=1} |\langle Tv, v\rangle|.$$

We need the following propositions:

Proposition 5. $|\langle Tv, v\rangle| \leq r(T)\|v\|^2$ *for all* $v \in V$.

Proof. If $v = 0$, then this is trivial. Otherwise, write

$$|\langle Tv, v\rangle| = \left|\left\langle T\left(\frac{v}{\|v\|}\right), \frac{v}{\|v\|}\right\rangle\right| \|v\|^2,$$

and by definition of $r(T)$ the desired inequality follows. □

Proposition 6. $r(T) \leq \|T\| \leq 4r(T)$.

Proof. Recall that

$$\|T\| = \sup_{\|v\|=\|w\|=1} |\langle Tv, w\rangle|$$

by Proposition 2. It follows immediately that $r(T) \leq \|T\|$. On the other hand, for all $v, w \in V$, we have

$$\begin{aligned}\langle Tv, w\rangle =& \frac{1}{4}\left(\langle T(v+w), v+w\rangle + i\langle T(v+iw), v+iw\rangle\right.\\ &\left. + i^2\langle T(v+i^2w), v+i^2w\rangle + i^3\langle T(v+i^3w), v+i^3w\rangle\right).\end{aligned}$$

Thus taking absolute value, we get, from the previous proposition, that

$$|\langle Tv, w\rangle| \leq \frac{1}{4}\left(r(T)\|v+w\|^2 + r(T)\|v+iw\|^2 + r(T)\|v+i^2w\|^2 + r(T)\|v+i^3w\|^2\right),$$

and taking sup over all $v, w \in V$ with $\|v\| = \|w\| = 1$, we get

$$\|T\| \leq \frac{1}{4}(r(T)2^2 + r(T)2^2 + r(T)2^2 + r(T)2^2) = 4r(T).$$

□

Proposition 7. $r(T^2) \leq r(T)^2$.

Proof. Notice that $r(cT) = cr(T)$ for all $c \geq 0$. Therefore, by replacing T with a multiple of T if necessary, we will without loss of generality assume that $r(T) = 1$. Then we have to prove $r(T^2) \leq 1$. The key is the following identity: for all $v \in V$,

$$\langle T^2v, v\rangle = \frac{1}{2}\left(\langle T(v+Tv), v+Tv\rangle - \langle T(v-Tv), v-Tv\rangle\right) - \|Tv\|^2.$$

[4]Again all results in this subsection holds, if V is a complex inner product space (possibly infinite dimensional), and if $T\colon V \to V$ is a *continuous* linear operator on V.

One can see this by simplifying the right hand side. Now the parallelogram law says that

$$\|Tv\|^2 = -\|v\|^2 + \frac{1}{2}(\|v+Tv\|^2 + \|v-Tv\|^2)$$

(which again one can check by expanding the right hand side). Thus substituting this into the previous equation, we get

$$\begin{aligned} \langle T^2 v, v\rangle - \|v\|^2 =& \frac{1}{2}\left(\langle T(v+Tv), v+Tv\rangle - \|v+Tv\|^2\right) \\ &+ \frac{1}{2}\left(\langle T(v-Tv), v-Tv\rangle - \|v-Tv\|^2\right). \end{aligned} \tag{2}$$

Suppose now $v_0 \in V$ is given. Let θ_0 be a real number such that $e^{2i\theta_0}\langle Tv_0, v_0\rangle = |\langle Tv_0, v_0\rangle|$. Let $T_0 = e^{i\theta_0}T$. Then apply (2) to T_0 and v_0, we get

$$\begin{aligned} |\langle T^2 v_0, v_0\rangle| - \|v_0\|^2 =& \langle T_0^2 v_0, v_0\rangle - \|v_0\|^2 \\ =& \frac{1}{2}\left(\langle T_0 v_1, v_1\rangle - \|v_1\|^2\right) + \frac{1}{2}\left(\langle T_0 v_2, v_2\rangle - \|v_2\|^2\right) \end{aligned}$$

where we define $v_1 = v_0 + T_0 v_0$ and $v_2 = v_0 - T_0 v_0$. Taking real parts of both sides, we get

$$\begin{aligned} |\langle T^2 v_0, v_0\rangle| - \|v_0\|^2 \leq& \frac{1}{2}\left(\mathrm{Re}\,\langle T_0 v_1, v_1\rangle - \|v_1\|^2\right) + \frac{1}{2}\left(\mathrm{Re}\,\langle T_0 v_2, v_2\rangle - \|v_2\|^2\right) \\ \leq& \frac{1}{2}\left(|\langle T_0 v_1, v_1\rangle| - \|v_1\|^2\right) + \frac{1}{2}\left(|\langle T_0 v_2, v_2\rangle| - \|v_2\|^2\right) \\ \leq& 0 \end{aligned}$$

since $r(T) = 1$ implies $r(T_0) = 1$. This proves $|\langle T^2 v_0, v_0\rangle| \leq \|v_0\|^2$ for any $v_0 \in V$. Thus $r(T^2) \leq 1$, as desired. □

Proposition 8. *If T is normal, i.e. if $TT^* = T^*T$, then $r(T) = \|T\|$. In particular, if T is self-adjoint, i.e. if $T^* = T$, then $r(T) = \|T\|$.*

Proof. By Proposition 6, we already have

$$r(T) \leq \|T\| \leq 4r(T);$$

this actually holds without having to assume that T is normal. On the other hand, suppose T is normal, i.e. $TT^* = T^*T$. We will get rid of the extra factor 4 in the previous inequality. This we do by going to higher powers of T. In fact, for any positive integer k, we have

$$\|T\|^{2^{k+1}} = (\|T\|^2)^{2^k} = \|T^*T\|^{2^k}$$

by Proposition 3. Now T^*T is self-adjoint, since $(T^*T)^* = T^*(T^*)^* = T^*T$. Thus by Proposition 4,

$$\|T\|^{2^{k+1}} = \|(T^*T)^{2^k}\|.$$

Since $TT^* = T^*T$, this is the same as $\|(T^*)^{2^k} T^{2^k}\|$. But $(T^*)^{2^k} = (T^{2^k})^*$, so by Proposition 4 again, we get

$$\|T\|^{2^{k+1}} = \|T^{2^k}\|^2.$$

Now apply Proposition 6: we get

$$\|T\|^{2^{k+1}} \leq \left(4r(T^{2^k})\right)^2.$$

But Proposition 7, applied k times, then implies

$$\|T\|^{2^{k+1}} \leq \left(4r(T)^{2^k}\right)^2 = 16r(T)^{2^{k+1}}.$$

Thus

$$\|T\| \leq 16^{2^{-(k+1)}} r(T)$$

for all $k \in \mathbb{N}$. Letting $k \to \infty$ we get $\|T\| \leq r(T)$, as desired. Thus we have the first assertion in the Proposition. The second assertion then follows easily, since every self-adjoint operator is normal. □

3.3. **Proof of (i) via the operator norm and the numerical radius.** We now come back to our original setting. Suppose V is a finite dimensional complex inner product space, and $T : V \to V$ is a self-adjoint linear operator on V. We will prove that T has an eigenvector in V. First, by Proposition 8, $\|T\| = r(T)$. But $r(T) = \sup_{\|v\|=1} F(v)$ where $F(v) := |\langle Tv, v\rangle|$ is a continuous real-valued function on the unit sphere $S := \{v \in V : \|v\| = 1\}$. Since V is finite dimensional, S is a compact set, and thus the supremum defining $r(T)$ is attained by a vector $v_1 \in S$. It follows that $\|v_1\| = 1$, and

$$\|T\| = r(T) = |\langle Tv_1, v_1\rangle| \leq \|Tv_1\|\|v_1\| \leq \|T\|,$$

where the first inequality is Cauchy-Schwarz, and the second inequality follows from the definition of $\|T\|$. But then the above chain of inequalities implies

$$|\langle Tv_1, v_1\rangle| = \|Tv_1\|\|v_1\|, \tag{3}$$

and by the equality case in the Cauchy-Schwarz, we have $Tv_1 = \lambda_1 v_1$ for some $\lambda_1 \in \mathbb{C}$. This shows v_1 is an eigenvector of T, and we have now a third proof the proof of (i) in the context of Theorem 2.

In fact the only place where we have used the self-adjointness of T in this subsection is when we invoked Proposition 8. Since one only needs T to be normal when applying Proposition 8, we have proved the more general fact that every normal linear operator $T : V \to V$ on a finite dimensional complex inner product space has an eigenvector in V.

4. Proof of Theorem 1

We now give the modifications that are necessary to prove Theorem 1. Thus in this section, unless otherwise stated, we assume that V is a real inner product space, and $T : V \to V$ is a symmetric linear operator. As in the proof of Theorem 2, we only need to prove the following:

(i) T has an eigenvector in V;
(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W \to W$ is symmetric.

The proof of (ii) in our current context is the same as that of the corresponding statement in the proof of Theorem 2. One just needs to keep in mind that if V is a real inner product space, then any eigenvalue of a symmetric linear operator

$T\colon V \to V$ is real (by definition). One then replaces any occurrence of T^* in the previous proof by T^t. We leave the details to the reader.

To prove (i) in our current context, we also give three different proofs.

First, we give a proof using the fundamental theorem of algebra, although this is probably not the most direct proof possible. If $\{u_1, \dots, u_n\}$ is a basis of V, we construct a complex inner product space[5] $\tilde{V}$ by letting $\tilde{V}$ be a complex vector space with basis $\{u_1, \dots, u_n\}$, and defining a complex inner product on $\tilde{V}$ by requiring

$$\left\langle \sum_{j=1}^n a_j u_j, \sum_{k=1}^n a_k u_k \right\rangle := \sum_{j=1}^n \sum_{k=1}^n a_j \overline{a_k} \langle u_j, u_k \rangle \quad \text{for all } a_1, \dots, a_n \in \mathbb{C}.$$

This definition of the complex inner product is independent of the choice of the basis $\{u_1, \dots, u_n\}$. Now extend $T\colon V \to V$ to a (complex) linear operator $\tilde{T}\colon \tilde{V} \to \tilde{V}$, by letting

$$\tilde{T}\left(\sum_{j=1}^n a_j u_j\right) = T\left(\sum_{j=1}^n b_j u_j\right) + iT\left(\sum_{k=1}^n c_j u_j\right),$$

where $a_1, \dots, a_n \in \mathbb{C}$, and b_j, c_j are the real and imaginary parts of a_j respectively. Again this extension is independent of the choice of $\{u_1, \dots, u_n\}$. One can then check that $\tilde{T}$ is self-adjoint on $\tilde{V}$; in fact $\tilde{T}^*$ is then the complex linear extension of $T^t\colon V \to V$ to $\tilde{V}$. Thus by the fundamental theorem of algebra, since $\tilde{V}$ is finite dimensional, $\tilde{T}$ has an eigenvalue $\lambda_1 \in \mathbb{C}$, say with eigenvector $\tilde{v_1} \in \tilde{V}$. But since $\tilde{T}$ is self-adjoint, we must have $\lambda_1 \in \mathbb{R}$. If $\tilde{v_1} = \sum_{j=1}^n \alpha_j u_j$, where $\alpha_j = \beta_j + i\gamma_j$ for some real numbers β_j and γ_j, then by multiplying $\tilde{v_1}$ by i if necessary, we may assume that some of the β_j's is not equal to zero. Then taking the real parts of the equation $\tilde{T}\tilde{v_1} = \lambda_1 \tilde{v_1}$, we get $Tv_1 = \lambda_1 v_1$, where $v_1 = \sum_{j=1}^n \beta_j v_j \in V$, and $v_1 \neq 0$. This shows that T has an eigenvector in V.

Next, we give an alternative proof of (i) using a variational method. Again let S be the unit sphere in V, i.e. $S = \{v \in V \colon \|v\| = 1\}$. It is a compact subset of V by the finite dimensionality of V. Now let $F(v) = \langle Tv, v \rangle$. This defines a smooth function from S to $\mathbb{R}$. Thus F achieves its maximum at some $v_1 \in S$. We can then show that v_1 is an eigenvector of T, pretty much in the same way that we had when we gave the second proof of (i) in the previous section. In fact, we then have

$$\left.\frac{d}{dt}\right|_{t=0} F\left(\frac{v_1 + tw}{\|v_1 + tw\|}\right) = 0,$$

which implies (as in the derivation of (1)) that

$$-2\langle v_1, w\rangle \langle Tv_1, v_1 \rangle + 2\langle Tv_1, w \rangle = 0.$$

Thus letting $\lambda_1 = \langle Tv_1, v_1 \rangle$, we get

$$\langle Tv_1 - \lambda_1 v_1, w \rangle = 0$$

for all $w \in V$. Thus $Tv_1 = \lambda_1 v_1$, and T has an eigenvector v_1.

[5] $\tilde{V}$ is really just the complexification of the real vector space V.

Finally, we give a proof of (i) using the operator norm and the numerical radius of T.

4.1. **Prerequisites.** If V is a finite dimensional[6] real inner product space, and $T\colon V \to V$ is a linear operator on V, the operator norm of T is still defined by

$$\|T\| = \sup_{\|v\|=1} \|Tv\|,$$

and the numerical radius of T is still defined by

$$r(T) = \sup_{\|v\|=1} |\langle Tv, v\rangle|.$$

It is still true that

$$\|T\| = \sup_{\|v\|=\|w\|=1} |\langle Tv, w\rangle|$$

and that $r(T) \leq \|T\|$; however, it is no longer true that $\|T\| \leq 4r(T)$. In fact $r(T)$ could be zero without having $\|T\| = 0$ (e.g. if T is a rotation in $\mathbb{R}^2$). Nevertheless,

Proposition 9. *If T is also symmetric, then we have*

$$r(T) = \|T\|.$$

Proof. This is because first

$$r(T) \leq \|T\|$$

trivially; second, by the symmetry of T, we have

$$\langle Tv, w\rangle = \frac{1}{4}(\langle T(v+w), v+w\rangle - \langle T(v-w), v-w\rangle)$$

for all $v, w \in V$, so

$$\|T\| \leq 2r(T);$$

in fact

$$\|T\| = \sup_{\|v\|=\|w\|=1} |\langle Tv, w\rangle| \leq \frac{1}{4}(r(T)2^2 + r(T)2^2) = 2r(T).$$

Also, by the symmetry of T, we claim that

$$r(T^2) \leq r(T)^2; \tag{4}$$

this is because if we go through the proof of Proposition 7, we still have the analog of (2), so if $r(T) = 1$, then for all $v \in V$, we have

$$\langle T^2 v, v\rangle - \|v\|^2 \leq 0,$$

i.e.

$$\langle T^2 v, v\rangle \leq \|v\|^2.$$

Now $\langle T^2 v, v\rangle$ is non-negative, thanks to the symmetry of T; thus the above says

$$|\langle T^2 v, v\rangle| \leq \|v\|^2$$

[6]Again in the case when V is infinite dimensional, the statements in this subsection remain true, as long as $T\colon V \to V$ is a *continuous* linear operator.

for all $v \in V$. This proves $r(T^2) \leq 1$, and proves our claim (4) whenever T is symmetric. Next one can still show, as before, that $\|T\|^{2^k} = \|T^{2^k}\|$ if T is symmetric. So putting these together, if T is symmetric, then

$$\|T\|^{2^k} = \|T^{2^k}\| \leq 2r(T^{2^k}) \leq 2(r(T))^{2^k},$$

which implies

$$\|T\| \leq 2^{2^{-k}} r(T) \quad \text{for all } k \in \mathbb{N},$$

or $\|T\| \leq r(T)$ as $k \to \infty$. Together with the reverse inequality we observed at the beginning, we get the desired claim. □

4.2. **Proof of (i) using the operator norm and numerical radius.** We can now give a third proof of (i) in our current context. In other words, by making use of the operator norm and the numerical radius, we will prove that T has an eigenvector in V when $T \colon V \to V$ is a symmetric linear operator on a finite dimensional real inner product space. The proof is in fact almost identical to that in subsection 3.3; the only difference is that since we are now in a real inner product space, when we invoke the identity case of the Cauchy-Schwarz, what we should conclude, from the analog of (3), is that $Tv_1 = \lambda_1 v_1$ for some $\lambda_1 \in \mathbb{R}$. Thus we still have the existence in V of an eigenvector of T, and we conclude the proof.

5. Proof of Theorem 3

In this section, suppose V is a finite dimensional complex inner product space, and $T \colon V \to V$ is a normal linear operator. We will prove that V has a basis that consists of eigenvectors of T, thereby completing the proof of Theorem 3.

Again the idea is to prove the following:

(i) T has an eigenvector in V;
(ii) If v_1 is an eigenvector of T and W is the orthogonal complement of v_1 in V, then W is T-invariant, and that $T|_W : W \to W$ is normal.

Once these are proved, the proof that (a) implies (b) in Theorem 3 can be completed as in the cases of Theorems 1 and 2. We leave that to the reader.

First we prove an important property about eigenvectors of normal operators.

Proposition 10. *Suppose $T \colon V \to V$ is normal. If v is an eigenvector of T with eigenvalue λ, then v is also an eigenvector of T^* with eigenvalue $\overline{\lambda}$, and conversely.*

Proof. By normality of T, we know that $T - \lambda I$ and $T^* - \overline{\lambda} I$ commutes for all $\lambda \in \mathbb{C}$. Thus for any $v \in V$ and any $\lambda \in \mathbb{C}$, we have

$$\langle (T - \lambda I)v, (T - \lambda I)v \rangle = 0 \Leftrightarrow \langle (T^* - \overline{\lambda} I)v, (T^* - \overline{\lambda} I)v \rangle = 0.$$

But this says

$$(T - \lambda I)v = 0 \Leftrightarrow (T^* - \overline{\lambda} I)v = 0.$$

So we have the desired assertion. □

We are now ready to prove statement (ii) above. Suppose again that $T: V \to V$ is a complex normal operator. If v_1 is an eigenvector of T, say with eigenvalue λ, let W be the orthogonal complement of v_1. Now if $w \in W$, we want to show that $Tw \in W$. In other words, we want to show $\langle Tw, v_1 \rangle = 0$. But

$$\langle Tw, v_1 \rangle = \langle w, T^* v_1 \rangle = \langle w, \overline{\lambda} v_1 \rangle = \lambda \langle w, v_1 \rangle = 0,$$

where the second equality follows from the previous proposition. Thus W is T-invariant. Similarly, one can show that W is T^*-invariant. Hence, if U is the restriction of T to W, then for all $w_1, w_2 \in W$, we have

$$\langle w_1, U^* w_2 \rangle = \langle U w_1, w_2 \rangle = \langle T w_1, w_2 \rangle = \langle w_1, T^* w_2 \rangle,$$

which shows $U^* w_2 = T^* w_2$ for all $w_2 \in W$. (Here one uses $T^* w_2 \in W$.) From normality of T, it then follows that $U^*U = UU^*$, and thus the restriction of T to W is still a normal operator.

Next, we turn to a proof of (i). In fact we will give three different proofs.

The most direct one is to use the fundamental theorem of algebra, as in the first proof of (i) in Theorem 2. We leave the details to the reader.

The second proof is one that makes use of the operator norm and the numerical radius. In fact we observed at the end of subsection 3.3 that the proof there proves (i) in our current context.

The third proof makes use of the following characterization of normal operators:

Proposition 11. *Suppose V is a finite dimensional complex inner product space. Given a linear operator $T: V \to V$, we define*

$$T_1 = \frac{1}{2}(T + T^*), \quad \textit{and} \quad T_2 = \frac{1}{2i}(T - T^*).$$

Then $T = T_1 + iT_2$, T_1 and T_2 are self-adjoint, and T is normal if and only if $T_1 T_2 = T_2 T_1$.

Sometimes the T_1 and T_2 defined here are called the real and imaginary parts of T respectively.

Proof. It is obvious that $T = T_1 + iT_2$, $T_1^* = T_1$ and $T_2^* = T_2$. It follows that

$$T^* = T_1 - iT_2.$$

Thus

$$T^*T = (T_1 - iT_2)(T_1 + iT_2) = T_1^2 + T_2^2 + i(T_1T_2 - T_2T_1),$$

while

$$TT^* = (T_1 + iT_2)(T_1 - iT_2) = T_1^2 + T_2^2 - i(T_1T_2 - T_2T_1).$$

Thus $TT^* = T^*T$ if and only if $T_1T_2 = T_2T_1$. □

Now by the proposition, T_1 and T_2 are self-adjoint operators. In particular, by Theorem 2, they are diagonalizable operators. Furthermore, the above proposition also says that T_1 and T_2 commutes with each other. Thus by the theorem about commuting diagonalizable operators, they are simultaneously diagonalizable. It follows that they have at least one common eigenvector, say v_0, in V. Then v_0 is an eigenvector of T, since $T = T_1 + iT_2$. This gives a third proof of (i) in our current context.

6. Schur's lemma

Many proofs above are rather analytic in nature. For instance, they make use of the derivative, or the supremum of smooth functions over compact sets. On the other hand, there is a more algebraic proof of these theorems. This relies on the splitting of the characteristic polynomial of a linear operator, and the key is the following Schur's lemma.

Theorem 4 (Schur's lemma)**.** *Suppose V is a finite dimensional inner product vector space over a field F, where $F = \mathbb{C}$ or $\mathbb{R}$. Suppose $T\colon V \to V$ is a linear operator. If the characteristic polynomial of T splits over F, then V has an ordered orthonormal basis β such that the matrix representation $[T]_\beta$ is upper triangular.*

Proof. The proof is by induction on the dimension of V. We will do it in the case when the scalar field $F = \mathbb{C}$; a small modification of the proof will give the case when $F = \mathbb{R}$. So suppose now $F = \mathbb{C}$. First, since the characteristic polynomial of T splits over F, T has an eigenvalue, say $\lambda \in F$. Then $T - \lambda I$ is not invertible, so so is $(T - \lambda I)^* = T^* - \overline{\lambda} I$. Thus $\overline{\lambda}$ is an eigenvalue of T^*, and thus T^* has an eigenvector in V, say v_1, with eigenvalue $\overline{\lambda}$. We may normalize it so that $\|v_1\| = 1$. Let W be the orthogonal complement of v_1 in V. Then W is T-invariant: in fact, if $w \in W$, then

$$\langle Tw, v_1\rangle = \langle w, T^* v_1\rangle = \langle w, \overline{\lambda} v_1\rangle = \lambda \langle w, v_1\rangle = 0.$$

Thus W is T-invariant. But let U be the restriction of T to W; then the characteristic polynomial of U divides the characteristic polynomial of T. So the characteristic polynomial of U also splits over F, and by induction hypothesis, since $\dim(W) < \dim(V)$, there is an ordered orthonormal basis γ of W such that the matrix representation of U with respect to this basis is upper triangular. It follows that $\beta := \gamma \cup \{v_1\}$ is an ordered orthonormal basis of V, and that the matrix representation of T with respect to β is upper triangular. This completes our induction. □

With this we can give yet another proof of the implication (a) implies (b) in the Theorems 1, 2 and 3. We illustrate this with the implication in Theorem 3.

Suppose V is a finite dimensional complex inner product space, and $T\colon V \to V$ is a linear operator. Then by the fundamental theorem of algebra, the characteristic polynomial of T splits over $\mathbb{C}$. Thus by Schur's lemma, V admits an ordered orthonormal basis β for with the matrix representation $[T]_\beta$ is upper triangular. We will denote this basis by $\beta = \{v_1, \dots, v_n\}$. Now suppose further that T is normal. Observe that v_1 is an eigenvector of T. Thus by Proposition 10, v_1 is also an eigenvector of T^*. It follows that the first column of $[T^*]_\beta$ is zero except for the first entry. But since β is an orthonormal basis, we have $[T^*]_\beta = ([T]_\beta)^*$. Thus the first row of $[T]_\beta$ is zero except for the first entry. This implies v_2 is an eigenvector of T as well; so by normality of T again, v_2 is an eigenvector of T^*. It follows that the second column of $[T^*]_\beta$ is zero except for the second entry; so the second row of $[T]_\beta$ is zero except in the second entry. Repeating, we see that the j-th row of $[T]_\beta$ is zero except for the j-th entry, for all $j = 1, \dots, n$; thus $[T]_\beta$ is diagonal. This proves that β is an orthonormal basis of V that consists of eigenvectors of T, and this proves the implication (a) implies (b) in Theorem 3 again.

An analogous argument can be used to prove the implication (a) implies (b) in Theorems 1 and 2. We leave the modifications to the interested reader.

7. Matrix versions of the main Theorems

Having proved our Theorems in Section 1, we state now the corresponding versions for matrices. First we recall some definitions.

An $n \times n$ real matrix P is said to be *orthogonal* if and only if $PP^t = P^tP = I$; here P^t is the transpose of the matrix P. It is known that if P is an $n \times n$ real matrix, then the following statements are equivalent:

(i) P is orthogonal;
(ii) $P^tP = I$;
(iii) the columns of P form an orthonormal basis of $\mathbb{R}^n$;
(iv) $PP^t = I$;
(v) the rows of P form an orthonormal basis of $\mathbb{R}^n$.

An $n \times n$ complex matrix U is said to be *unitary* if and only if $U^*U = UU^* = I$; here U^* is the conjugate transpose (or adjoint) of the matrix U. It is known that if U is an $n \times n$ complex matrix, then the following statements are equivalent:

(i) U is unitary;
(ii) $U^*U = I$;
(iii) the columns of U form an orthogonal basis of $\mathbb{C}^n$;
(iv) $UU^* = I$;
(v) the rows of U form an orthogonal basis of $\mathbb{C}^n$.

An $n \times n$ real matrix A is said to be *symmetric* if and only if $A^t = A$. An $n \times n$ complex matrix A is said to be *Hermitian* (or *self-adjoint*) if and only if $A^* = A$. An $n \times n$ complex matrix A is said to be *normal* if and only if $AA^* = A^*A = I$.

It follows that if one uses the standard inner products on $\mathbb{R}^n$ and $\mathbb{C}^n$, then A is symmetric if and only if the linear operator $x \mapsto Ax$ is a symmetric operator on $\mathbb{R}^n$, and A is self-adjoint (resp. normal) if and only if the linear operator $z \mapsto Az$ is a self-adjoint (resp. normal) operator on $\mathbb{C}^n$.

Thus we have the following theorems. The first is a characterization of (real) symmetric matrices:

Theorem 5. *Suppose A is an $n \times n$ real matrix. Then the following are equivalent:*

(a) A is symmetric;
(b) $\mathbb{R}^n$ has an orthonormal basis that consists of eigenvectors of A;
(c) there exists an orthogonal matrix P, and a diagonal matrix D, such that $A = PDP^{-1}$.

Note that it follows from the last statement of the above theorem that any symmetric matrix is diagonalizable over $\mathbb{R}$. Also, one should observe that in last statement of the above theorem, one could also have written

$$A = PDP^t,$$

since $P^t = P^{-1}$ when P is an orthogonal matrix.

Similarly, we have the following characterization of Hermitian matrices:

Theorem 6. *Suppose A is an $n \times n$ complex matrix. Then the following are equivalent:*

(a) A is Hermitian;
(b) $\mathbb{C}^n$ has an orthogonal basis that consists of eigenvectors of A, and all eigenvalues of A are real;
(c) there exists an unitary matrix U, and a diagonal matrix D with real entries, such that $A = UDU^{-1}$.

Finally we have the following characterization of normal matrices:

Theorem 7. *Suppose A is an $n \times n$ complex matrix. Then the following are equivalent:*

(a) A is normal;
(b) $\mathbb{C}^n$ has an orthogonal basis that consists of eigenvectors of A;
(c) there exists an unitary matrix U, and a diagonal matrix D, such that $A = UDU^{-1}$.

Again, it follows from the last statement of the two theorems above that any self-adjoint matrix is diagonalizable over $\mathbb{C}$, and so is any normal matrix. Also, one should observe that in last statement of the two theorems above, one could also have written

$$A = UDU^*,$$

since $U^* = U^{-1}$ when U is a unitary matrix.