

Complex inner products (6.7 supplement)

The definition of inner product given in section 6.7 of Lay is not useful for complex vector spaces because no nonzero complex vector space has such an inner product. If it did, pick any vector $\mathbf{u} \neq \mathbf{0}$ and then $0 < \langle \mathbf{u}, \mathbf{u} \rangle$. But also

$$0 < \langle i\mathbf{u}, i\mathbf{u} \rangle = i\langle \mathbf{u}, i\mathbf{u} \rangle = i^2\langle \mathbf{u}, \mathbf{u} \rangle = -\langle \mathbf{u}, \mathbf{u} \rangle < 0$$

which is a contradiction. Following is an altered definition which will work for complex vector spaces. Recall that if $z = a + bi$ then $\bar{z} = a - bi$ is the complex conjugate of z . We alter the definition of inner product by taking complex conjugate sometimes.

Definition A Hermitian inner product on a complex vector space V is a function that, to each pair of vectors \mathbf{u} and \mathbf{v} in V , associates a complex number $\langle \mathbf{u}, \mathbf{v} \rangle$ and satisfies the following axioms, for all $\mathbf{u}, \mathbf{v}, \mathbf{w}$ in V and all scalars c :

1. $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\langle \mathbf{v}, \mathbf{u} \rangle}$.
2. $\langle \mathbf{u} + \mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{w} \rangle + \langle \mathbf{v}, \mathbf{w} \rangle$ and $\langle \mathbf{u}, \mathbf{v} + \mathbf{w} \rangle = \langle \mathbf{u}, \mathbf{v} \rangle + \langle \mathbf{u}, \mathbf{w} \rangle$.
3. $\langle c\mathbf{u}, \mathbf{v} \rangle = \bar{c}\langle \mathbf{u}, \mathbf{v} \rangle$ and $\langle \mathbf{u}, c\mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$.¹
4. $\langle \mathbf{u}, \mathbf{u} \rangle$ is a nonnegative real number and $\langle \mathbf{u}, \mathbf{u} \rangle = 0$ if and only if $\mathbf{u} = \mathbf{0}$.

The vector space \mathbb{C}^n has a standard inner product, $\langle \mathbf{u}, \mathbf{v} \rangle = \mathbf{u}^* \mathbf{v}$. Recall $\mathbf{u}^* = \overline{\mathbf{u}}^T$ so another formula is $\langle \mathbf{u}, \mathbf{v} \rangle = \overline{\mathbf{u}}^T \mathbf{v}$. So for example

$$\langle [1+i, 2-i], [3-2i, 1+i] \rangle = (1-i)(3-2i) + (2+i)(1+i) = 1-5i+1+3i = 2-2i.$$

You can compute this in Matlab as `dot([1+i, 2-i], [3-2i, 1+i])`.

Another example is the complex vector space $C_{\mathbb{C}}[a, b]$ of complex valued continuous functions with domain $[a, b]$. So any vector in $C_{\mathbb{C}}[a, b]$ is of the form $f(t) + ig(t)$ where f and g are in $C[a, b]$. Examples are $t^2 + t^3i$ and $e^{5ti} = \cos(5t) + \sin(5t)i$. We can define a Hermitian inner product on $C_{\mathbb{C}}[a, b]$ by $\langle u, v \rangle = \int_a^b \overline{u(t)}v(t) dt$. So for example in $C_{\mathbb{C}}[0, 2\pi]$

$$\langle e^{kti}, e^{\ell ti} \rangle = \int_0^{2\pi} \overline{e^{kti}} e^{\ell ti} dt = \int_0^{2\pi} e^{(\ell-k)ti} dt = \left. \frac{e^{(\ell-k)ti}}{(\ell-k)i} \right|_0^{2\pi} = 0$$

if $\ell \neq k$. (Note that $\overline{e^{ai}} = e^{-ai}$.)

Just as for inner products, the length of a vector \mathbf{u} is defined as $\|\mathbf{u}\| = \sqrt{\langle \mathbf{u}, \mathbf{u} \rangle}$. The angle θ between two vectors \mathbf{v} and \mathbf{u} is defined by $\|\mathbf{u}\| \|\mathbf{v}\| \cos \theta = \Re(\langle \mathbf{u}, \mathbf{v} \rangle)$, where $\Re(\langle \mathbf{u}, \mathbf{v} \rangle)$ denotes the real part of the Hermitian inner product. For example, to find the angle between $[3 i]^T$ and $[2 + i \ 1 - i]^T$ in \mathbb{C}^2 , we have:

$$\begin{aligned} \|[3 i]^T\| &= \sqrt{3 \cdot 3 + (-i) \cdot i} = \sqrt{10} \\ \|[2 + i \ 1 - i]^T\| &= \sqrt{(2-i) \cdot (2+i) + (1+i) \cdot (1-i)} = \sqrt{2^2 + 1^2 + 1^2 + 1^2} = \sqrt{7} \\ \langle [3 i]^T, [2 + i \ 1 - i]^T \rangle &= 3 \cdot (2+i) + (-i) \cdot (1-i) = 6 + 3i - i - 1 = 5 + 2i \\ \theta &= \cos^{-1}(5/(\sqrt{10}\sqrt{7})) \approx .93 \text{ radians} \end{aligned}$$

I promised to prove Cauchy-Schwarz for a Hermitian inner product for anyone interested. Here goes. If $z = a + bi$ is a complex number we define $|z| = \sqrt{\bar{z}z} = \sqrt{a^2 + b^2}$. First, we show that $|\langle \mathbf{v}, \mathbf{u} \rangle| \leq 1$ for all unit vectors \mathbf{v} and \mathbf{u} . To see this suppose, \mathbf{v} and \mathbf{u} are unit vectors, so $\|\mathbf{v}\| = 1$ and $\|\mathbf{u}\| = 1$. Then for any complex number t we have:

$$0 \leq \|\mathbf{u} - t\mathbf{v}\|^2 = \langle \mathbf{u} - t\mathbf{v}, \mathbf{u} - t\mathbf{v} \rangle = \langle \mathbf{u}, \mathbf{u} \rangle + \langle \mathbf{u}, -t\mathbf{v} \rangle + \langle -t\mathbf{v}, \mathbf{u} \rangle + \langle -t\mathbf{v}, -t\mathbf{v} \rangle$$

¹ There is some disagreement on this axiom. Some authors say $\langle c\mathbf{u}, \mathbf{v} \rangle = c\langle \mathbf{u}, \mathbf{v} \rangle$ and $\langle \mathbf{u}, c\mathbf{v} \rangle = \bar{c}\langle \mathbf{u}, \mathbf{v} \rangle$. We choose our form somewhat arbitrarily. But, for example, it is the definition used by Matlab where the Hermitian inner product is calculated by `dot(u, v)`.

$$= \langle \mathbf{u}, \mathbf{u} \rangle - t\langle \mathbf{u}, \mathbf{v} \rangle - \bar{t}\langle \mathbf{v}, \mathbf{u} \rangle + t\bar{t}\langle \mathbf{v}, \mathbf{v} \rangle = 1 - t\langle \mathbf{u}, \mathbf{v} \rangle - \bar{t}\langle \mathbf{v}, \mathbf{u} \rangle + t\bar{t}$$

Let $t = \langle \mathbf{v}, \mathbf{u} \rangle$. Then by property 1 of the Hermitian inner product, we know that $\bar{t} = \langle \mathbf{u}, \mathbf{v} \rangle$. Then

$$0 \leq 1 - t\langle \mathbf{u}, \mathbf{v} \rangle - \bar{t}\langle \mathbf{v}, \mathbf{u} \rangle + t\bar{t} = 1 - t\bar{t} = 1 - |t|^2$$

So $1 \geq |t| = |\langle \mathbf{v}, \mathbf{u} \rangle|$ as we desired. Now suppose \mathbf{v} and \mathbf{u} are any vectors. Then $\mathbf{v}/\|\mathbf{v}\|$ and $\mathbf{u}/\|\mathbf{u}\|$ are unit vectors, so we just showed that

$$1 \geq |\langle \mathbf{v}/\|\mathbf{v}\|, \mathbf{u}/\|\mathbf{u}\| \rangle| = |\langle \mathbf{v}, \mathbf{u} \rangle| / (\|\mathbf{v}\| \cdot \|\mathbf{u}\|)$$

Consequently, $\|\mathbf{v}\| \cdot \|\mathbf{u}\| \geq |\langle \mathbf{v}, \mathbf{u} \rangle|$.

Hermitian and Unitary matrices (supplement to 7.1)

There is a complex version of orthogonal matrices. A complex square matrix U is called **unitary** if $U^* = U^{-1}$. Equivalently, the columns of U form an orthonormal set (using the standard Hermitian inner product on \mathbb{C}^n). Any orthogonal matrix is unitary.

Likewise, there is a complex version of symmetric matrices. A complex square matrix A is called **Hermitian** if $A^* = A$. What this boils down to is that the diagonal entries of A are all real, and the above diagonal entries are the conjugates of the corresponding below diagonal entries. Any symmetric matrix is Hermitian. The spectral theorem applies to Hermitian matrices and indeed it is most easily proven for Hermitian matrices.

Since Lay does not provide a proof of the spectral theorem I will sketch a proof below.

Theorem 1. *If $T: V \rightarrow V$ is a linear transformation of a nonzero finite dimensional complex vector space V then T has an eigenvector.*

Proof: By taking coordinates with respect to a basis of V , we may as well suppose that $V = \mathbb{C}^n$ and T is multiplication by a complex matrix A . Also $n > 0$ since V is nonzero. Then the characteristic polynomial of A has at least one root, so A has at least one eigenvector, which corresponds to an eigenvector of T . ■

Theorem 2. *The product of two unitary matrices is unitary.*

Proof: Suppose Q and S are unitary, so $Q^{-1} = Q^*$ and $S^{-1} = S^*$. Then $(QS)^* = S^*Q^* = S^{-1}Q^{-1} = (QS)^{-1}$ so QS is unitary. ■

Theorem 3. (Schur Lemma) *If A is any square complex matrix then there is an upper triangular complex matrix U and a unitary matrix S so that $A = SUS^* = SUS^{-1}$.*

Proof: Let q_1 be an eigenvector of A , which we may suppose has unit length. By the Gram-Schmidt process we may choose q'_i so that $\{q_1, q'_2, \dots, q'_n\}$ is an orthonormal basis. Let $Q_0 = [q_1 q'_2 \cdots q'_n]$, then Q_0 is unitary and $Q_0^* A Q_0 = \begin{bmatrix} \lambda_1 & * \\ 0 & A_1 \end{bmatrix}$ for some $(n-1) \times (n-1)$ matrix A_1 . Likewise, we may find a unitary $(n-1) \times (n-1)$

matrix Q_1 so that $Q_1^* A_1 Q_1 = \begin{bmatrix} \lambda_2 & * \\ 0 & A_2 \end{bmatrix}$. Then if $S_1 = Q_0 \begin{bmatrix} 1 & 0 \\ 0 & Q_1 \end{bmatrix}$ we have $S_1^* A S_1 = \begin{bmatrix} \lambda_1 & * & * \\ 0 & \lambda_2 & * \\ 0 & 0 & A_2 \end{bmatrix}$.

Note that S_1 is unitary by Theorem 2. Now continue in this fashion, letting $S_k = S_{k-1} \begin{bmatrix} I_k & 0 \\ 0 & Q_k \end{bmatrix}$, and we see that $U = S_n^* A S_n$ is upper triangular. Letting $S = S_n$ we see that $A = SUS^*$. ■

We are now prepared to prove the Spectral theorem. Note that if A is a real symmetric matrix then it is Hermitian so the following result implies Theorem 3 on page 452 of Lay.

Theorem 4. (Spectral theorem) *If A is an $n \times n$ Hermitian matrix then:*

- A has n real eigenvalues, counting multiplicities.*
- The dimension of the eigenspace for each eigenvalue λ equals the multiplicity of λ as a root of the characteristic polynomial of A .*
- The eigenspaces are mutually orthogonal, in the sense that eigenvectors corresponding to different eigenvalues are orthogonal.*
- There is a unitary matrix S and a real diagonal matrix D so that $A = SDS^* = SDS^{-1}$.*
- If A is real then we may choose S in d) to be real (orthogonal).*

Proof: By Theorem 3 there is a unitary matrix S and an upper triangular U so that $A = SUS^*$. But $A = A^*$ so

$$U = S^*AS = S^*A^*S = (S^*AS)^* = U^*$$

but then U must be diagonal since U and U^* are both upper triangular. Moreover, if λ is a diagonal entry of U , then $\bar{\lambda}$ is the corresponding entry of U^* which means $\lambda = \bar{\lambda}$, which means λ is real. So setting $D = U$ we have shown d). Since S diagonalizes A we know that the diagonal entries of D are the eigenvalues of A and the columns of S are the eigenvectors. So we have shown a) and b) (see theorem 5 page 320 and Theorem 7 page 324). Note c) follows since the columns of S are orthogonal. If A is real we might need to choose a new S since the S we constructed in Theorem 3 might not be real, unless we chose it very carefully. For each eigenvalue λ_j of A , choose an orthonormal (real) basis \mathcal{B}_j of the eigenspace corresponding to λ_j . Let the columns of S be all these basis vectors in $\mathcal{B}_1, \dots, \mathcal{B}_k$. By b) this will make n columns. Then S^*AS is diagonal and by part c), S is orthogonal.

I will now give an alternative proof which does not use the Schur Lemma. As above, let \mathcal{B}_j be an orthonormal basis of the λ_j eigenspace. By Theorem 1, page 450 we know that $\mathcal{B}_1, \mathcal{B}_2, \dots, \mathcal{B}_k$ form an orthonormal set $\{q_1, \dots, q_\ell\}$. Complete it to an orthonormal basis $\{q_1, \dots, q_n\}$ of \mathbb{C}^n . Note that if $\ell = n$ we are done since we have an orthonormal basis of eigenvectors, so suppose that $\ell < n$. Let W be the span of $\{q_{\ell+1}, \dots, q_n\}$. Note that vectors in W are perpendicular to all eigenspaces of A . If $i > \ell$ and $j \leq \ell$ then

$$\langle Aq_i, q_j \rangle = (Aq_i)^* q_j = q_i^* A^* q_j = q_i^* A q_j = q_i^* \lambda q_j = \lambda \langle q_i, q_j \rangle = 0$$

consequently Aq_i is perpendicular to all q_j for $j \leq \ell$ which means Aq_i is in W . Consequently, $A:W \rightarrow W$. So by Theorem 1, A has an eigenvector in W . But this contradicts the fact that all vectors in W are perpendicular to all eigenspaces of W , a nonzero vector can't be perpendicular to itself. So we conclude that in fact $\ell = n$ and $W = \{0\}$ so we are done. ■

Finally we characterize which matrices can be diagonalized by a unitary matrix. We say a matrix A is **normal** if $AA^* = A^*A$.

Theorem 5. *A matrix A is diagonalizable with a unitary matrix if and only if A is normal. In other words:*

- a) *If A is normal there is a unitary matrix S so that S^*AS is diagonal.*
- b) *If there is a unitary matrix S so that S^*AS is diagonal then A is normal.*

Proof: Suppose A is normal. By Theorem 3 there is a unitary matrix S and an upper triangular U so that $A = SUS^*$. Then

$$UU^* = S^*AS(S^*AS)^* = S^*ASS^*A^*S = S^*AA^*S = S^*A^*AS = S^*A^*SS^*AS = U^*U$$

But if we let u_{ij} denote the ij -th entry of U then the upper left entry of U^*U is $u_{11}\bar{u}_{11} = |u_{11}|^2$ but the upper left entry of UU^* is

$$u_{11}\bar{u}_{11} + u_{12}\bar{u}_{12} + \dots + u_{1n}\bar{u}_{1n} = |u_{11}|^2 + |u_{12}|^2 + \dots + |u_{1n}|^2$$

Since this equals $|u_{11}|^2$ and all summands are nonnegative real numbers we must have $u_{12} = u_{13} = \dots = u_{1n} = 0$. Similarly, looking at the second diagonal entry we see that $u_{2j} = 0$ for all $j > 2$. Continuing in this way we see that U must be diagonal. So we have shown that if A is normal, then it is diagonalizable with a unitary matrix.

Now suppose that A is any matrix so that there is a unitary matrix S so that $S^*AS = D$ is diagonal. Note $DD^* = D^*D$. Then

$$AA^* = SDS^*(SDS^*)^* = SDS^*SD^*S^* = SDD^*S^* = SD^*DS^* = SD^*S^*SDS^* = A^*A$$

Consequently, A is normal. ■

Examples of normal matrices are Hermitian matrices ($A = A^*$), skew Hermitian matrices ($A = -A^*$) and unitary matrices ($A^* = A^{-1}$) so all such matrices are diagonalizable.