1 Some examples

1. The functions below are examples of inner products:

   (a). $V = \mathbb{C}([0,1]) = \{f : [0,1] \to \mathbb{C} \text{ continuous}\}$.
   $< f, g >= \int_0^1 f\overline{g}$

   (b). $V = M_n(\mathbb{C}), < A, B >= \text{tr}(AB^*)$, where $B^* = \overline{B^t}$.
   Proof: The conditions of the inner products can be established as below:
   • $< A+B, C >= \text{tr}((A+B)C^*) = \text{tr}(AC^* + BC^*) = \text{tr}(AC^*) + \text{tr}(BC^*) =< A+C, B+C >$.
   • for any constant c, $< cA, B >= \text{tr}(c(AB^*)) = c\text{tr}(AB^*) = c < A, B >$.
   • $< A, B >= \text{tr}(AB^*) = \text{tr}(A\overline{B^t}) = \sum(A\overline{B^t})_{ii} = \sum A_{ij}\overline{B^t_{ji}} = \sum A_{ij}\overline{B_{ij}}$, $\forall 1 \leq i, j \leq n$. Similarly, $< B, A >= \sum B_{ij}\overline{A_{ij}}$. Now for $a, b \in \mathbb{C}$ we have $\overline{a} + \overline{b} = \overline{a+b}$, $\overline{ab} = \overline{a}\overline{b}$ and $\overline{\overline{a}} = a$. Therefore $\overline{a\overline{b}} = \overline{a}\overline{\overline{b}} = \overline{a}b$. This gives $A_{ij}\overline{B_{ij}} = \overline{B_{ij}\overline{A_{ij}}}$ and therefore $< A, B >= A_{ij}\overline{B_{ij}} = \overline{B_{ij}\overline{A_{ij}}} = \overline{< B, A >}$.
   • From above, $< A, A >= \sum A_{ij}\overline{A_{ij}} = \sum ||A_{ij}||^2$. This is obviously nonnegative, and it is zero if and only if all $||A_{ij}||$'s are zero, meaning that A_{ij} must be itself a zero (i.e. a zero vector).

2. In assignment 1 problem 1, we have seen that the pairing isn't an inner product because there exists nonzero vector $\boldsymbol{x}$ satisfying $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0$. We now show that the pairing $\boldsymbol{x}^t \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \overline{\boldsymbol{y}}$ satisfies all other properties.

   Notice that, if $\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$ and $\boldsymbol{y} = \begin{pmatrix} y_1 \\ y_2 \end{pmatrix}$ then

   $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \boldsymbol{x}^t A \overline{\boldsymbol{y}} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} 1 & i \\ -i & 1 \end{pmatrix} \begin{pmatrix} \overline{y_1} \\ \overline{y_2} \end{pmatrix} = \begin{pmatrix} x_1 & x_2 \end{pmatrix} \begin{pmatrix} \overline{y_1} + i\overline{y_2} \\ -i\overline{y_1} + \overline{y_2} \end{pmatrix} = \begin{pmatrix} x_1(\overline{y_1} + i\overline{y_2}) + x_2(-i\overline{y_1} + \overline{y_2}) \end{pmatrix}.$$

   We establish the following:
   • $\langle \boldsymbol{x} + \boldsymbol{y}, \boldsymbol{z} \rangle = \boldsymbol{x} + \boldsymbol{y}^t A\overline{\boldsymbol{z}} = (\boldsymbol{x}^t + \boldsymbol{y}^t)A\overline{\boldsymbol{z}} = \boldsymbol{x}^t A\overline{\boldsymbol{z}} + \boldsymbol{y}^t A\overline{\boldsymbol{z}} = \langle \boldsymbol{x}, \boldsymbol{z} \rangle + \langle \boldsymbol{y}, \boldsymbol{z} \rangle$.
   • For any constant c, $\langle \boldsymbol{cx}, \boldsymbol{y} \rangle = (c\boldsymbol{x}^t)A\overline{\boldsymbol{y}} = c(\boldsymbol{x}^t A\overline{\boldsymbol{y}}) = c\langle \boldsymbol{x}, \boldsymbol{y} \rangle$.
   • Before proving $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}$, we need the following properties about complex numbers: for any complex numbers a and b, we have $\overline{a} + \overline{b} = \overline{a+b}$; for any complex numbers a and b, $\overline{a} \cdot \overline{b} = \overline{ab}$. Therefore,

   $$\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \begin{pmatrix} x_1(\overline{y_1} + i\overline{y_2}) + x_2(-i\overline{y_1} + \overline{y_2}) \end{pmatrix}, \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \begin{pmatrix} y_1(\overline{x_1} + i\overline{x_2}) + y_2(-i\overline{x_1} + \overline{x_2}) \end{pmatrix}.$$

   We have $x_1\overline{y_1} = \overline{\overline{x_1}\,\overline{\overline{y_1}}} = \overline{\overline{x_1}y_1}$, and similarly $x_2\overline{y_2} = \overline{\overline{x_2}\,\overline{\overline{y_2}}} = \overline{\overline{x_2}y_2}$. In addition, $i(x_1\overline{y_2} - x_2\overline{y_1}) = i(\overline{\overline{x_1}y_2 - \overline{x_2}y_1}) = -i(\overline{\overline{x_2}y_1 - \overline{x_1}y_2}) = \overline{i}(\overline{\overline{x_2}y_1 - \overline{x_1}y_2}) = \overline{i(\overline{x_2}y_1 - \overline{x_1}y_2)}$. Therefore,

   $x_1(\overline{y_1}+i\overline{y_2})+x_2(-i\overline{y_1}+\overline{y_2}) = x_1\overline{y_1}+x_2\overline{y_2}+i(x_1\overline{y_2}-x_2\overline{y_1}) = \overline{\overline{x_1}y_1}+\overline{\overline{x_2}y_2}+\overline{i(\overline{x_2}y_1 - \overline{x_1}y_2)} = \overline{\overline{x_1}y_1 + \overline{x_2}y_2 + i(\overline{x_2}y_1 - \overline{x_1}y_2)}$

   $= \overline{y_1(\overline{x_1} + i\overline{x_2}) + y_2(-i\overline{x_1} + \overline{x_2}}$, establishing the claim.
   • Now $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \begin{pmatrix} x_1(\overline{x_1} + i\overline{x_2}) + x_2(-i\overline{x_1} + \overline{x_2}) \end{pmatrix} = (x_1\overline{x_1} + x_2\overline{x_2} + i(x_1\overline{x_2} - x_2\overline{x_1}) = |x_1|^2 + |x_2|^2 + ix_1\overline{x_2} + (-i)\overline{x_1}x_2 = |x_1|^2 + |x_2|^2 + ix_1\overline{x_2} + \overline{\overline{i}\,\overline{x_1}\,\overline{\overline{x_2}}} = |x_1|^2 + |x_2|^2 + ix_1\overline{x_2} + \overline{ix_1\overline{x_2}} = |x_1|^2 + |x_2|^2 + 2Re(ix_1\overline{x_2})$, because $a + \overline{a} = 2Re(a)$. Now, $|2Re(ix_1\overline{x_2})| \leq |2(ix_1\overline{x_2})| \leq 2|x_1x_2|$ so $-2|x_1x_2| \leq |2Re(ix_1\overline{x_2})| \leq 2|x_1x_2|$, so $|x_1|^2 + |x_2|^2 + 2Re(ix_1\overline{x_2}) \geq |x_1|^2 + |x_2|^2 - 2|x_1||x_2| = (|x_1| - |x_2|)^2$, so the pairing is always nonnegative. Notice, however, it could happen that this quantity is indeed 0 even with both x_1, x_2 nonzero.

2 Proofs of identities

1. Given basis $\{\vec{w_1}, \cdots, \vec{w_n}\}$ of an inner product space, prove that the the set of vectors $\{\vec{v_1}, \cdots, \vec{v_n}\}$ defined as $\vec{v_1} = \vec{w_1}$ and

   $$\vec{v_k} = \vec{w_k} - \sum_{i=1}^{k-1} \frac{\langle \vec{w_k}, \vec{v_i} \rangle}{||\vec{v_i}||^2} \vec{v_i}, \qquad \forall\, k \in [2, n]$$

   is an orthogonal basis.

Proof: (Credits to textbook and prof). First, we prove that $\langle\vec{i},\vec{j} = 0$, $\forall i \neq j$. We also proeed by inducting on n. Base case where $n = 1$ is trivial. Suppose the claim holds for $n = 1, 2, \cdots k-1$ for some k, we have: for any $j < k$,

$$\langle\vec{v_k},\vec{v_j}\rangle = \langle\vec{w_k} - \sum_{i=1}^{k-1}\frac{\langle\vec{w_k},\vec{v_i}\rangle}{||\vec{v_i}||^2}\vec{v_i},\vec{v_j}\rangle = \langle\vec{w_k},\vec{v_j}\rangle - \langle\sum_{i=1}^{k-1}\frac{\langle\vec{w_k},\vec{v_i}\rangle}{||\vec{v_i}||^2}\vec{v_i},\vec{v_j}\rangle = \langle\vec{w_k},\vec{v_j}\rangle - \sum_{i=1}^{k-1}\frac{\langle\vec{w_k},\vec{v_i}\rangle}{||\vec{v_i}||^2}\langle\vec{v_i},\vec{v_j}\rangle$$

$$= \langle\vec{w_k},\vec{v_j}\rangle - \frac{\langle\vec{w_k},\vec{v_j}\rangle}{||\vec{v_j}||^2}\langle\vec{v_j},\vec{v_j}\rangle = \langle\vec{w_k},\vec{v_j}\rangle - \langle\vec{w_k},\vec{v_j}\rangle = 0,$$

justifying the claim. (By induction hypothesis, $\langle\vec{i},\vec{j}\rangle = 0$ for any $i < j < k$.

Next, notice that none of the vectors $\vec{v_i}$ can be zero; each of the vectors $\vec{v_k}$ can be written as the linear combination of $\vec{w_1}, \cdots, \vec{w_k}$, with the coefficient of $\vec{w_k}$ being 1. Since $\vec{w_1}, \cdots, \vec{w_k}$ are linearly independent, the claim follows.

Finally, in class we have seen that a set of nonzero orthogonal vectors must be linearly independent. Since the set of vectors $\{\vec{v_1}, \cdots, \vec{v_n}\}$ has n elements and are linearly independent, this set is also a basis. The conclusion follows.

2. Given a finite dimensional inner-product space V and let W be its subspace with orthonormal basis $\{\vec{w_1}, \cdots, \vec{w_k}\}$. Then for each $\vec{v} \in V$ there exists a unique $\vec{w} \in W$ and $\vec{w'} \in W^\perp$ satisfying $\vec{w} + \vec{w'} = \vec{v}$, given by the following formula:

$$\vec{w} = \sum_{i=1}^{k}\frac{\langle\vec{v},\vec{w_i}\rangle}{||\vec{w_i}||^2}\vec{w_i}, \qquad \vec{w'} = \vec{v} - \vec{w}.$$

Proof: since a subspace (or a vector space, in general) is closed under addition, $\vec{w}$ described above is in W. To show that $\vec{w'} \in W^\perp$, we notice the following for all $j \in [1, n]$:

$$\langle\vec{w'},\vec{w_j}\rangle = \langle\vec{v} - \sum_{i=1}^{k}\frac{\langle\vec{v},\vec{w_i}\rangle}{||\vec{w_i}||^2}\vec{w_i},\vec{w_j}\rangle = \langle\vec{v},\vec{w_j}\rangle - \langle\sum_{i=1}^{k}\frac{\langle\vec{v},\vec{w_i}\rangle}{||\vec{w_i}||^2}\vec{w_i},\vec{w_j}\rangle = \langle\vec{v},\vec{w_j}\rangle - \sum_{i=1}^{k}\frac{\langle\vec{v},\vec{w_i}\rangle}{||\vec{w_i}||^2}\langle\vec{w_i},\vec{w_j}\rangle = \langle\vec{v},\vec{w_j}\rangle - \langle\vec{v},\vec{w_j}\rangle = 0,$$

because $\langle\vec{w_i},\vec{w_j}\rangle$ vanishes whenever $i \neq j$, and $\frac{\langle\vec{v},\vec{w_j}\rangle}{||\vec{w_j}||^2}\langle\vec{w_j},\vec{w_j}\rangle = \langle\vec{v},\vec{w_j}\rangle$.

To show that the numbers $\vec{w}$ and $\vec{w'}$ are unique, suppose that there exists $\vec{w_1}, \vec{w_2} \in W$ and $\vec{w_1'}, \vec{w_2'} \in W^\perp$ satisfying $\vec{w_1} + \vec{w_1'} = \vec{w_2} + \vec{w_2'}$. Now, $\vec{w_1} - \vec{w_2} \in W$ and $\vec{w_1'} - \vec{w_2'} = -(\vec{w_1} - \vec{w_2}) \in W^\perp$, which means the vector $\vec{w_1} - \vec{w_2}$ is in both W and $W^\perp$ (the product of any vector in W and any scalar constant is also in W). Notice, however, that this means $||\vec{w_1} - \vec{w_2}|| = 0$ by the definition of W and $W^\perp$, so $\vec{w_1} - \vec{w_2} = 0$ or $\vec{w_1} = \vec{w_2}$, showing that such pair of numbers must be unique.

3. Let V be a finite dimensional transformation. Then for each transformation $T : V \to V$ there is a unique tranformation T^* satisfying $\langle T(\vec{x}),\vec{y}\rangle = \langle\vec{x},T^*(\vec{y})\rangle$ for all $\vec{x},\vec{y} \in V$.

Proof: Let n be the dimension of V, and denote $\{\vec{v_1}, \cdots, \vec{v_n}\}$ by an orthonormal basis of V. We use the fact that each linear transformation is uniquely determined by the values of $T(\vec{v_1}), \cdots, T(\vec{v})n)$. That is, for each n-tuples of vectors $\{\vec{w_1}. \cdots, \vec{w_n}\}$ there is a unique linear transformation T such that $T(\vec{v_i}) = \vec{w_i}$. Suppose that numbers $a_{ij} \in \mathbb{C}, 1 \leq i, j \leq n$ are such that $T(\vec{v_i}) = \sum_{i=1}^{n} a_{ij}\vec{v_j}$, we have, for each i, k, $\langle T(\vec{v_i}),\vec{v_k}\rangle = \langle\sum_{i=1}^{n} a_{ij}\vec{v_j},\vec{v_k}\rangle = a_{ik}$. Suppose that there is a linear transformation T^* such that $\langle T(\vec{x}),\vec{y}\rangle = \langle\vec{x},T^*(\vec{y})\rangle$ for all $\vec{x},\vec{y} \in V$. Let b_{ij} be numbers such that $T^*(\vec{v_i}) = \sum_{i=1}^{n} b_{ij}\vec{v_j}$ then we have $a_{ik} = \langle T(\vec{v_i}),\vec{v_k}\rangle = \langle(\vec{v_i}),T^*(\vec{v_k})\rangle = \overline{\langle T^*(\vec{v_k}),\vec{v_i}\rangle} = \overline{\langle\sum_{i=1}^{n} b_{kj}\vec{v_j},\vec{v_i}\rangle} = \overline{a_{ki}}$, therefore we must have $T^*(\vec{v_i}) = \sum_{i=1}^{n} b_{ij}\vec{v_j} = T^*(\vec{v_i}) = \sum_{i=1}^{n} \overline{b_{ji}}\vec{v_j}$. This uniquely defines T^*.

Conversely, let T^* be as defined, given T. From above we already have the relation $\langle T(\vec{v}_i), \vec{v}_k\rangle = \langle(\vec{v}_i), T^*(\vec{v}_k)\rangle$ for each pair of vectors in our orthonormal basis. Let $\vec{x} = \sum_{i=1}^{n} x_i\vec{v}_i$ and $\vec{y} = \sum_{i=1}^{n} y_i\vec{v}_i$ then we have

$$\langle T(\vec{x}), \vec{y}\rangle = \langle T(\sum_{i=1}^{n} x_i\vec{v}_i), \sum_{i=1}^{n} y_i\vec{v}_i\rangle = \langle \sum_{i=1}^{n} x_i T(\vec{v}_i), \sum_{i=1}^{n} y_i\vec{v}_i\rangle = \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\overline{y_j}\langle T(\vec{v}_i), \vec{v}_j\rangle$$

$$= \sum_{i=1}^{n}\sum_{j=1}^{n} x_i\overline{y_j}\langle \vec{v}_i, T^*(\vec{v}_j)\rangle = \langle \sum_{i=1}^{n} x_i\vec{v}_i, \sum_{j=1}^{n} y_j T^*(\vec{v}_j)\rangle = \langle \sum_{i=1}^{n} x_i\vec{v}_i, T^*(\sum_{j=1}^{n} y_j\vec{v}_j)\rangle = \langle \vec{x}, T^*(\vec{y})\rangle$$

4. Let $A = [T]_\beta$ for some orthonormal basis β is a finite dimensional space V. Then $[T]_\beta^* = [T^*]_\beta$.

   Proof: Let our orthonormal basis be $\{\vec{v}_1, \cdots, \vec{v}_n\}$. This proof relies on the following fact: $A_{ij} = \langle T(\vec{v}_j), \vec{v}_i\rangle$. This is because for each j, $[T\vec{v}_j]_\beta = [T]_\beta[\vec{v}_j]_\beta = \mathrm{Col}_j(A)$ so $\langle T(\vec{v}_j), \vec{v}_i\rangle = \sum_{i=1}^{n} A_{kj}\langle \vec{v}_k, \vec{v}_i\rangle = A_{ij}$, as desired. Thus for each i, j we have $([T]_\beta^*)_{ij} = (A^*)_{ij} = \overline{A_{ij}^t} = \overline{A_{ji}} = \overline{\langle T(\vec{v}_i), \vec{v}_j\rangle} = \overline{\langle \vec{v}_i, T^*(\vec{v}_j)\rangle} = \langle T^*(\vec{v}_j), \vec{v}_i\rangle = ([T^*]_\beta)_{ij}$.

5. A transformation $T : V \to V$ (over complex numbers) is orthogornally diagonalizable if and only if $TT^* = T^*T$.

   Proof. (Creds: both our prof and the textbook). Suppose that $[T]_\beta$ is diagonal for some orthonormal basis β. Then the following holds:

$$[TT^*]_\beta = [T]_\beta[T^*]_\beta = [T]_\beta[T]_\beta^* = [T]_\beta^*[T]_\beta = [T^*]_\beta[T]_\beta = [T^*T]_\beta$$

   Notice the implicit use of the fact that $[T]_\beta^*$ is diagonal, every two diagonal matrices commute, and that $[T]_\beta^* = [T^*]_\beta$ because β is orthonormal.

   To prove the converse, we need the following Schur's lemma: for each transformation T whose characteristic polynomial splits there exists an orthonormal basis β such that $[T]_\beta$ is upper triangular. To prove this, let's do induction on n, the dimension of T. Base case $n = 1$ is obvious. The inductive step relies on the following fundamental theorem of algebra. Every complex polynomial (the characteristic polynomial, in partiular), has a complex root. Thus there exists a $z \neq 0$ such that $T(z) = \lambda z$. Therefore for any y we have:

$$0 = \langle (T - \lambda I)z, y\rangle = \langle z, (T - \lambda I)^* y\rangle = \langle z, (T^* - \overline{\lambda} I)y\rangle$$

   Therefore $z \in [im(T^* - \overline{\lambda} I)]^\perp$, and the rank-nullity theorem suggests the existence of an x such that $x \in \ker(T^* - \overline{\lambda} I)$, which means $T^* I = \overline{\lambda} x$ (which suugests that if λ is an eigenvector of T then $\overline{\lambda}$ is an eigenvector of T^*.) This means, the subspace $W = \{x\}$ is T^*-invariant. Since for each $g \in W^\perp$ we have: $\langle T(g), x\rangle = \langle g, T^* x\rangle = \langle g, \overline{\lambda} x\rangle = \lambda\langle g, x\rangle = 0$, $W^\perp$ is T invariant. In addition, $\dim(W^\perp) = \dim(V) - 1 = n - 1$. The characteristic polynomial of $T_{W^\perp}$ divides that of T, and hence splits. This allows us to use our inductive hypothesis on the existence of an orthonormal basis $\beta' = \{\vec{v}_1, \cdots, \vec{v}_{n-1}\}$ such that $[T_{W^\perp}]_{\beta'}$ is upper triangular. Combining this with our new vector x we get $\beta = \{\vec{v}_1, \cdots, \vec{v}_{n-1}, \vec{x}\}$, another set of orthonormal basis (because $\{\vec{v}_1, \cdots, \vec{v}_{n-1} \in W^\perp\}$) and the resulting matrix $[T]_\beta$ will be upper triangular. □

   Having proven the lemma, we proceed to our main problem. Assume that $TT^* = T^*T$ as defined in our problem, and let β be orthonormal such that $A = [T]_\beta$ is upper triangular. Now, $A = [T]_\beta^* = [T^*]_\beta$ so $AA* = A * A$. We will prove this directly by equating the coefficients. By the upper triangularity of A we have $A_{ij} = 0$ for any $i > j$. Also we have:

$$\sum_{k=1}^{n} A_{ik}\overline{A_{jk}} = \sum_{k=1}^{n} A_{ik}A_{kj}^* = (AA^*)_{ij} = (A^*A)_{ij} = \sum_{k=1}^{n} A_{ik}^* A_{kj} = \sum_{k=1}^{n} \overline{A_{ki}} A_{kj}$$

Suppose that for some $p \geq 0$, $A_{ij} = 0$ for any $i \neq j$ and $i \leq p$. ($k = 0$ is the case where we haven't proven anything). Now, letting $i = j = p + 1$ we have:

$$\sum_{k=p+1}^{n} |A_{(p+1)k}| = \sum_{k=1}^{n} |A_{(p+1)k}| = \sum_{k=1}^{n} A_{(p+1)k}\overline{A_{(p+1)k}} = \sum_{k=1}^{n} \overline{A_{k(p+1)}} A_{k(p+1)} = \sum_{k=1}^{n} |A_{k(p+1)}| = \sum_{k=1}^{p+1} |A_{k(p+1)}|$$

By the inductive hypothesis, the last quantity is actually equal to $|A_{(p+1)(p+1)}|$. This forces $\sum_{k=p+2}^{n} |A_{(p+1)k}| = 0$, and by the positive definiteness of absolute value we have $A_{(p+1)k} = 0$ for all $k \neq p+1$. This finishes the proof that A is diagonal. Q.E.D.

6. Every eigenvector of a self-adjoint transformation is real.

   Proof: Since $T = T^*$, T is normal and hence diagonalizable in some orthonormal basis β (allowing complex eigenvectors and eigenvalues instead of real). Now T_β is diagonal with eigenvector $\lambda_i = T_{ii}$, but $\lambda_i = T_{ii} = \overline{T^*_{ii}} = \overline{\lambda_i}$ so λ_i is real.

7. Let $T : \mathbb{R}^n \to \mathbb{R}^n$ to be an orthogonal transformation. Then there exists a basis β such that T_β is real and block diagonal with each block having size at most 2.

   Proof: Since T is orthogonal, it is diagonalizable in some basis α, although the eigenvectors or eigenvalues might be complex numbers. Now for each real matrix A, and $A\vec{v} = \lambda\vec{v}$ for some v we have $\overline{A} = A$ so $A\overline{\vec{v}} = \overline{A\vec{v}} = \overline{\lambda\vec{v}} = \overline{\lambda}\overline{\vec{v}}$, where $\overline{\vec{v}}$ is the "coordinate-wise conjugate" of $\vec{v}$. Therefore, the eigenvalues and eigenvectors of T come in pairs (if complex). (Notice that $\vec{v}$ can be "stand-alone" it it's real).

   Now we rearrange the basis α to make it $\{\vec{v}_1, \overline{\vec{v}_1}, \vec{v}_3, \overline{\vec{v}_3}, \cdots, \vec{v}_{2k-1}, \overline{\vec{v}_{2k-1}}, \vec{v}_{2k+1}, \cdots, \vec{v}_n\}$; the first $2k$ of which are complex conjugate pairs and the last $n - 2k$ are real. We claim that

   $$\beta = \{\vec{v}_1 + \overline{\vec{v}_1}, i(\vec{v}_1 - \overline{\vec{v}_1}), \cdots, \vec{v}_{2k-1} + \overline{\vec{v}_{2k-1}}, i(\vec{v}_{2k-1} - \overline{\vec{v}_{2k-1}}), \vec{v}_{2k+1}, \cdots, \vec{v}_n\}$$

   will have $[T]_\beta$ in the form we want. First, notice that β is a real basis (proof skipped :P); second, the entries responsible for $\vec{v}_{2k+1}, \cdots, \vec{v}_n$ vanish except on the diagonals, and the diagonal entries are real eigenvalues. Finally, for each $\vec{v}_i + \overline{\vec{v}_i}$ and $i(\vec{v}_i - \overline{\vec{v}_i})$, denote W_i be the subspace spanned by $\{\vec{v}_i, \overline{\vec{v}_i}\}$. Since $\vec{v}_i$ and $\overline{\vec{v}_i}$ are the eigenvectors, T is W_i invariant, and so the entries of T_β respomsible for these two are block diagonal with size two. Finally, these block diagonal entries are also real (resembling 2×2 orthogonal matrices of rotations and reflections), because the members $\vec{v}_i + \overline{\vec{v}_i}$ and $i(\vec{v}_i - \overline{\vec{v}_i})$ are real. This conclude the proof.

8. Let $T : V \to V$ be a projection (in a real space). Then the following are equivalent:

   - T is orthogonal projection.
   - $\ker(T) = \mathrm{im}(T)^\perp$
   - $T = T^*$.

   Proof: We first prove the equivalence of the first two conditions. The fact that T is orthogonal projection means that there exists an othonormal nonzero vectors $W = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k\}$ such that $T(\vec{x}) = \sum_{i=1}^{k} \langle \vec{x}, \vec{v}_i \rangle \vec{v}_i$. Obviously $\mathrm{im}(T) = W$ and $\ker T = W^\perp$ since $T(\vec{x}) = 0$ iff $\langle \vec{x}, \vec{v}_i \rangle$ for all $i \in [1, k]$. Conversely, suppose that $\ker(T) = \mathrm{im}(T)^\perp$. Let $X = \{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k\}$ be an orthonormal basis of the $\mathrm{im}(T)$, and $W = \{\vec{w}_1, \cdots, \vec{w}_m\}$ be an orthonormal basis of $\ker(T)$. Then $\langle \vec{v}_i, \vec{w}_j \rangle = 0$, so it's not hard to prove that vectors in W and X are linearly independent of each other. By rank-nullity theorem, $X \cup W$ is an orthonormal basis of V. Since T is a projections, $T(\vec{v}_i) = \vec{v}_i$ and by the definition of null space $T(\vec{w}_i) = 0$. Thus $T\left(\sum_{i=1}^{k} a_i\vec{v}_i + \sum_{j=1}^{m} b_j\vec{w}_j\right) = \sum_{i=1}^{k} a_i T(\vec{v}_i) + \sum_{j=1}^{m} b_j T(\vec{w}_j) = \sum_{i=1}^{k} a_i\vec{v}_i = \sum_{i=1}^{k} \langle \sum_{j=1}^{k} a_j\vec{v}_j, \vec{v}_i \rangle \vec{v}_i$, hence an orthogonal projection from T onto $\mathrm{span}\{\vec{v}_1, \vec{v}_2, \cdots, \vec{v}_k\}$.

For the equivalence of the first and the third fact, we first show (3) implies (1). Now, $T = T^*$ so it's normal (and hence orthonormally diagonalizable). Let β an orthonormal basis whose members are eigenvectors of T. Then, from the fact that $T^2 = T$ we have $\lambda^2 = \lambda$ for all eigenvalues λ, hene $\lambda \in \{0, 1\}$. Now split the basis int two parts: $W : \{\vec{x} \in \beta, \lambda = 1\}$ and $X : \{\vec{x} \in \beta, \lambda = 0\}$. We now see that T is an orthogonal projection w.r.t. span(W). The relation (1) implies (3) is not that hard: indeed, if T is an orthogonal projection w.r.t. W for some subspace W or V, then W, then $W^\perp$ is a null space of T. Now let β be the union of the bases of W and $W^\perp$, then β is itself a basis of V. This means T_β is diagonal, with entry 1 at cell corresponding to W and 0 at cell corresponding to $W^\perp$, which is evidently self-adjoint.

9. We have seen the two definition of pseudoinverse in class:

   - If $A = U\Sigma V^*$ is the singular value decomposition of matrix A, then $A^\dagger = V\Sigma^\dagger U^*$ where $\Sigma^\dagger$ is the replacement of nonzero values in Σ with the reciprocals.
   - If T is any linear map, then $T^\dagger$ will give 0 for input in $(\text{im}(T))^\perp$, and give $T^{-1}|_{(\ker T)^\perp}$ for input in $\text{im}(T)$.

   They are equal.

   Proof: Let $T : P \to W$ be any linear map, and denote B_P and B_W as the bases of P and W that are used in the singular value decomposition of T. Denote $[T]_S^R = U\Sigma V^*$ as the singular value decomposition of $[T]_S^R$, and we want to prove that $[T^\dagger]_R^S = V\Sigma^\dagger U^*$. where S is any basis of P and R any basis of W. Denote $\vec{v}_1, \cdots, \vec{v}_n$ be the vectors in B_P and $\vec{w}_1, \cdots, \vec{w}_m$ be the vectors in B_W. First we need the following cheat fact: B_P (or V) contains column vectors that are either in $\ker(T)$ or $\ker(T)^\perp$, and B_W (or U) contains columns that are either in $\text{im}(T)$ or $\text{im}(T)^\perp$. To prove this, let r be the number of nonzero entries in Σ (and $\Sigma^\dagger$), we show that:

   - $\vec{v}_i \in \ker(T)$ if $i > r$ and $\in \ker(T)^\perp$ otherwise.
   - $\vec{w}_i \in \text{im}(T)$ if $i < r$ and $\in \text{im}(T)^\perp$ otherwise.

   Notice that $\text{im}(T) = \text{span}(\{T(\vec{v}_1, \cdots, T(\vec{v})_n\}) = \text{span}(\{\sigma_1\vec{w}_1, \cdots, \sigma_r\vec{w}_r, 0, 0, \cdots, 0\})$ so $\vec{v}_i$ are in $\ker(T)$ iff $i > r$. For $i \le r$, since B_P is orthogonal, the vectors are in $\ker(T)^\perp$. Meanwhile for $\vec{w}_i$, we have seen that they are in $\text{im}(T)$ if $i \le r$, and again since rank $(\text{im}(T)) = r$ and since B_W is also orthogonal, $\vec{w}_i \in \text{im}(T)^\perp$ for $i > r$. Therefore we must have $T^\dagger(\vec{w}_i) = 0$ for $i > r$, and since $T(\vec{v}_i) = \sigma(\vec{w}_i)$ for $i \le r$, we must have $T^\dagger(\vec{w}_i) = \sigma^{-1}\vec{v}_i$ for such i. To finish the proof on the equality when being feed into matrices, it suffices to consider just the basis B_W. Notice also by the definition of U, V we have $U^*[\vec{w}_i]_R^{B_W} = e_i$, with e_i the i-th element in the standard coordinate (since U is a change of basis matrix from B_W to R). Now there are two cases. For any $i > r$ we have

   $$V\Sigma^\dagger U^*[\vec{w}_i]_R = V\Sigma^\dagger e_i = V(0) = 0$$

   and similarly for $i \le r$ we have

   $$V\Sigma^\dagger U^*[(\vec{w}_i]_R = V\Sigma^\dagger e_i = V\sigma_i^{-1}e_i = \sigma_i^{-1}(\text{Col}_V(i)) = [\sigma(\vec{v}_i)]_S$$

   This gives us the desired equality. Ps: we could have replaced R and S with B_W and B_V, but that sounds cheating because all the matrices will be diagonal. Aha.

10. (Cayley-Hamilton Theorem) For finite-dimensional linear transformation $T : V \to V$, let p be its characteristic polynomial. Then $p(T)$ is the zero tranformation.

    Proof: We shall show that for each $\vec{v} \in \mathbb{C}^n$ (where $n \times n$ is the dimension of T) we have $p(A)(\vec{v}) = 0$. W.L.O.G. let $\vec{v} \neq 0$. Now consider the subspace W of $\mathbb{C}^n$ spanned by $\{T^i(\vec{v}) : i \ge 0\}$. Since $\mathbb{C}^n$ is finite dimensional, there must exists k such that $\{T^i(\vec{v}) : 0 \le i \le k\}$ is linearly dependent. We denote k (notation abuse alert! But don't care :P) as the minimum of such index such that the set is linearly dependent, then $\{T^i(\vec{v}) : 0 \le i \le k-1\}$ is linearly independent. We claim that this $\{T^i(\vec{v}) : 0 \le i \le k-1\}$ is a basis of W. It suffices to prove that $m \ge k \to T^m(\vec{v}) \in \text{span}(\{T^i(\vec{v}) : 0 \le i \le k-1\})$. Indeed, if $T^m(\vec{v}) = \sum_{i=0}^{k-1} a_i T^i(\vec{v})$, then $T^{m+1}(\vec{v}) = \sum_{i=0}^{k-1} a_i T(T^i(\vec{v})) = \sum_{i=1}^{k} a_i T^i(\vec{v})$ but we have assume that $T^k(\vec{v})$ is in the span of the set, so $T^{m+1}(\vec{v})$ is also in the span of the set. The conclusion follows from inductive hypothesis.

Now, We know that T restricted to W is W invariant, and $[T|_W]$ in our basis has the following form:

$$\begin{pmatrix} 0 & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & 0 & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & 0 & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{k-1} \end{pmatrix}$$

We need the following: the characteristic polynomial in the form:

$$\begin{vmatrix} -\lambda & 0 & 0 & \cdots & 0 & -a_0 \\ 1 & -\lambda & 0 & \cdots & 0 & -a_1 \\ 0 & 1 & -\lambda & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -a_{k-1} - \lambda \end{vmatrix}$$

Doesn't seem easy to start from, so we induce on k to claim that this is $q(\lambda) = (-1)^k \sum_{i=0}^{k} a_i \lambda^i$ (with $a_k = 1$ for ease of computation). Base case $k = 1$ we have $-\lambda - a_0 = (-1)(\lambda + a_0)$. Now suppose that the claim is true for $k - 1$ for some $k \geq 2$, then the characteristic polynomial is actually

$$-\lambda \begin{vmatrix} -\lambda & 0 & \cdots & 0 & -a_1 \\ 1 & -\lambda & \cdots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -a_{k-1} - \lambda \end{vmatrix} + (-1)^{k-1}(-a_0) \begin{vmatrix} 1 & -\lambda & 0 & \cdots & 0 \\ 0 & 1 & -\lambda & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{vmatrix}$$

By induction hypothesis the first term is given by $(-\lambda)(-1)^{k-1} \sum_{i=1}^{k} a_i \lambda^i$ and since it's not hard to see that the second matrix has determinant 1, then second term is actually $(-1)^k a_0$. Thus we have $(-\lambda)(-1)^{k-1} \sum_{i=1}^{k} a_i \lambda^i + (-1)^k a_0 = (-1)^k \sum_{i=0}^{k} a_i \lambda^i$ as desired. The matrix also implies that $T^k(\vec{v}) = -(\sum_{i=0}^{k-1} a_i T^i(\vec{v}))$. Therefore, $(-1)^k q(T)(\vec{v}) = \sum_{i=0}^{k} a_i T^i(\vec{v})) = \sum_{i=0}^{k-1} a_i T^i(\vec{v})) - (\sum_{i=0}^{k-1} a_i T^i(\vec{v})) = 0$, because $a_k = 1$ by our convention. This completes our proof.

Finally, extend the basis of W to form a basis of V and since T is W invariant, when T is written in the full basis of V we get $\begin{pmatrix} A & B \\ 0 & C \end{pmatrix}$ and so the characteristic polynomial would be $\begin{vmatrix} A - \lambda I & B \\ 0 & C - \lambda I \end{vmatrix}$, which is $\det(A - \lambda I)\det(C - \lambda I)$. However, $q(\lambda) = \det(A - \lambda I)$ so q divides p and since $q(T)(\vec{v}) = 0$, we must have $p(T)(\vec{v}) = 0$ too.

Corollary. If T is nilpotent, the $T^n = 0$ if n is the dimension of T.

11. For any transformation T of $V \to V$ over $\mathbb{C}$, denote K_λ as the generalized eigenspace of an eigenvalue λ. Then V is isomorphic to the direct sums of all the generalized eigenspace of T.

Proof: Let $(x - a_1)^{m_1} \cdot (x - a_k)^{m_k}$ be the characteristic polynomial of T, then $(T - a_1 I)^{m_1} \cdot (T - a_k I)^{m_k} = 0$ by Cayley-Hamilton theorem. Let n be the dimension of V, which is also $m_1 + m_2 + \cdots + m_k$. This means that $\dim(\ker((T-a_1I)^{m_1}))+\cdots+\dim(\ker((T-a_kI)^{m_k})) \geq n$. Now, let's show that the kernels are disjoint. Suppose that $\vec{v}$ satisfies $(T-a_iI)^c(\vec{v})=0$ and $(T-a_jI)^d(\vec{v})=0$. Since $(T-a_iI)^c$ and $(T-a_jI)^d$ has no roots in common, there exists constants polynomials p and q such that $p(T)(T-a_iI)^c+q(T)(T-a_jI)^d=I$, so adding them up yields $I(\vec{v})=0$, so $\vec{v}=0$. This, in turns, means that, $\dim(\ker((T-a_1I)^{m_1}))+\cdots+\dim(\ker((T-a_kI)^{m_k})) \leq n$ (recall the assumption that the kernel is always a subspace of V). Therefore $\dim(\ker((T-a_1I)^{m_1}))+\cdots+\dim(\ker((T-a_kI)^{m_k}))=n$, and by the fact that each of the kernel is disjoint, we can conclude that V is isomorphic to the direct sum of each of these generalized eigenspaces.