MA1112: Linear Algebra II

Dr. Vladimir Dotsenko (Vlad)

Lecture 21

The main slogan about Hermitian spaces is that the definition is designed in a way that allows us to re-prove most of what we know for Euclidean vector spaces. In particular, similarly to how one defines orthogonal and orthonormal systems in Euclidean spaces, we say that a system of vectors $e_1, \ldots, e_k$ of a Hermitian space V is orthogonal, if it consists of nonzero vectors, which are pairwise orthogonal: $(e_i, e_j) = 0$ for $i \neq j$. An orthogonal system is said to be orthonormal, if $(e_i, e_i) = 1$ for all i. There is an obvious version of Gram–Schmidt orthogonalisation procedure for Hermitian scalar products; it ensures that every Hermitian vector space has an orthonormal basis. Also, one can define orthogonal complements of subspaces, and prove their properties.

Note that a basis $e_1, \ldots, e_n$ of V is orthonormal if and only if

$$(x_1e_1 + \ldots + x_ne_n, y_1e_1 + \ldots + y_ne_n) = x_1\overline{y}_1 + \ldots + x_n\overline{y}_n,$$

so choosing an orthonormal basis identifies V with $\mathbb{C}^n$ equipped with the "standard" Hermitian structure.

Definition 1. Let φ be a linear transformation of a Hermitian vector space V. The *adjoint* linear transformation $\varphi^\dagger$ is uniquely determined by the formula

$$(\varphi v_1, v_2) = (v_1, \varphi^\dagger v_2)$$

that should be valid for all v_1 and v_2. (Indeed, substituting elements from an orthonormal basis for v_1 recovers all coordinates of $\varphi^\dagger v_2$).

If we work with a particular orthonormal basis $e_1, \ldots, e_n$ of a Hermitian vector space V, so that the linear transformation φ is represented by a matrix B relative to this basis, the linear transformation $\varphi^\dagger$ is represented by the matrix $\overline{B^T}$.

Lemma 1. *For every linear transformation φ, we have $(\varphi^\dagger)^\dagger = \varphi$.*

Proof. Indeed, we have

$$(v_1, (\varphi^\dagger)^\dagger(v_2)) = (\varphi^\dagger(v_1), v_2) = \overline{(v_2, \varphi^\dagger(v_1))} = \overline{(\varphi(v_2), v_1)} = (v_1, \varphi(v_2)),$$

from which it follows that

$$(\varphi^\dagger)^\dagger(v_2) = \varphi(v_2),$$

and it implies

$$(\varphi^\dagger)^\dagger = \varphi.$$

□

Definition 2. A linear transformation φ of a Hermitian vector space is said to be *symmetric*, or *Hermitian*, if $(\varphi(v_1), v_2) = (v_1, \varphi(v_2))$ for all v_1, v_2, in other words, if $\varphi^\dagger = \varphi$. A linear transformation is said to be *unitary*, if $(\varphi(v_1), \varphi(v_2)) = (v_1, v_2)$, in other words, if $\varphi^\dagger = \varphi^{-1}$. A linear transformation φ is said to be *normal* if $\varphi\varphi^\dagger = \varphi^\dagger\varphi$. (In particular, symmetric and unitary transformations are normal.)

Theorem 1. *A normal linear transformation admits an orthonormal basis of eigenvectors.*

Proof. We know that two commuting linear transformations have a common eigenvector, so there exists a vector v for which $\varphi(v) = cv$, and $\varphi^\dagger(v) = c'v$. We note that

$$c(v,v) = (cv,v) = (\varphi(v),v) = (v,\varphi^\dagger(v)) = (v,c'v) = \overline{c'}(v,v),$$

so $c = \overline{c'}$, although this will not be crucial for us. What is crucial is that for Hermitian scalar products one can define orthogonal complements etc., and it is easy to see that the orthogonal complement of v is an invariant subspace of both φ and $\varphi^\dagger$: if $w \in \operatorname{span}(v)^\perp$, then

$$(\varphi(w),v) = (w,\varphi^\dagger(v)) = (w,c'v) = \overline{c'}(w,v) = 0,$$
$$(\varphi^\dagger(w),v) = (w,(\varphi^\dagger)^\dagger(v)) = (w,\varphi(v)) = (w,cv) = c(w,v) = 0.$$

This allows us to proceed by induction on dimension of V: we found an eigenvector, and by the induction hypothesis we have a basis of eigenvectors in the orthogonal complement of that eigenvector. □

Theorem 2. *Every Hermitian linear transformation admits an orthonormal basis of eigenvectors. Eigenvalues of a symmetric linear transformation are real.*

Proof. We proved that every normal linear transformation admits an orthonormal basis of eigenvectors, so the first part follows. Moreover, we established in that proof that if v is a common eigenvector of φ and $\varphi^\dagger$, then the corresponding eigenvalues are complex conjugates of each other. Applying this to $\varphi = \varphi^\dagger$, we conclude that those eigenvalues are real. □

Remark 1. This theorem can be used to deduce the theorem on symmetric matrices we proved before, but the proof will invoke complex numbers. The proof we discussed before is a bit more consistent with the context of that theorem.

Outline application of linear algebra to face recognition

To conclude, I shall outline how linear algebra can be used to recognise faces in images. For simplicity, we shall assume that we have 80 grayscale passport photos, say 100×100 pixels, and we would like, once uploaded a new photo, to see if there are any existing photos that look like this one.

Grayscale images are represented by their *intensity vectors*: for each of 100×100 pixels, we record a number from 0 to 255 where 0 is black, 255 is white, and we have different shades of gray in between.

We regard individual pixels as random variables, so we have 10,000 random variables $\xi_1, \ldots, \xi_{10,000}$ and, for now, 80 different measurements for each ξ_i: $\xi_i^{(1)}, \ldots, \xi_i^{(80)}$. Assuming that the 80 faces are equidistributed, we can compute the covariance matrix

$$(\xi_i^{(1)} - E\xi_i^{(1)})(\xi_j^{(1)} - E\xi_j^{(1)}) + \cdots + (\xi_i^{(80)} - E\xi_i^{(80)})(\xi_j^{(80)} - E\xi_j^{(80)}).$$

Note that this is a symmetric matrix of a bilinear form, rather than of a linear map; similar to what we considered recently. We would like to change coordinates so that the covariance matrix becomes diagonal. This corresponds to finding a basis of uncorrelated (independent) random variables, which is really good because we identify principal face directions.

Let us look at our problem from another angle. We have 80 face vectors $v_1, \dots, v_{80}$ of size 36,000. We may compute their average

$$v := \frac{1}{80}(v_1 + \cdots + v_{80}),$$

the "mean face", and subtract it from all the given vectors, getting vectors $w_1, \dots, w_{80}$ centered at zero. If U is the $10,000 \times 80$-matrix made of these vectors, then the covariance matrix is equal to UU^T. This is a symmetric matrix of size 10,000, very hard to deal with. But, there is a very important property that we shall utilize: the nonzero eigenvalues of UU^T and U^TU are the same. How to show that? If $U^TUv = \lambda v$ for $\lambda \neq 0$, then $UU^TUv = \lambda Uv$, so Uv is an eigenvector of UU^T with eigenvalue λ. So instead of working with the matrix UU^T of size 36,000, we can work with the matrix U^TU of size 80 (up to eigenvalues 0 which are not supposed to contain important information about distinctive facial features). Eigenvectors of the latter should be multiplied by U to get eigenvectors $z_1, \dots, z_k$ of UU^T.

Now we can easily project every new face on the face space:

$$u \mapsto (u, z_1)z_1 + \cdots (u, z_k)z_k,$$

and search for closest one among the projection of the sample images. In practice, this can be improved in many ways, but it offers at least some flavour of how things work.