Summary of week 8 (Lectures 22, 23 and 24)

This week we completed our discussion of Chapter 5 of [VST].

Recall that if V and W are inner product spaces then a linear map $T\colon V \to W$ is called an *isometry* if it preserves lengths; that is, T is an isometry if and only if

$$\|T(x)\| = \|x\| \qquad \text{for all } x \in V.$$

It was shown at the end of Week 7 that this is equivalent to

$$\langle T(x), T(y)\rangle = \langle x, y\rangle \qquad \text{for all } x,\ y \in V.$$

In other words, T is an isometry if and only if it preserves inner products.

We have seen that for every linear map $T\colon \mathbb{R}^n \to \mathbb{R}^n$ there is a matrix $A \in \operatorname{Mat}(n \times n, \mathbb{R})$ such that

$$T(x) = Ax \qquad \text{for all } x \in \mathbb{R}^n.$$

It turns out that T is an isometry if and only if ${}^{\mathrm{t}}\!AA = I$, the identity matrix.

To prove this, suppose first of all that ${}^{\mathrm{t}}\!AA = I$. Then for all x, $y \in \mathbb{R}^n$,

$${}^{\mathrm{t}}x\, y = {}^{\mathrm{t}}x\, Iy = {}^{\mathrm{t}}x\, {}^{\mathrm{t}}\!AAy,$$

and so

$$T(x) \cdot T(y) = (Ax) \cdot (Ay) = {}^{\mathrm{t}}(Ax)(Ay) = {}^{\mathrm{t}}x\, {}^{\mathrm{t}}\!AAy = {}^{\mathrm{t}}x\, y = x \cdot y.$$

Thus T preserves the dot product and is therefore an isometry on $\mathbb{R}^n$.

Conversely, suppose that T is an isometry. Then $T(x) \cdot T(y) = x \cdot y$ for all x, $y \in \mathbb{R}^n$, and so

$${}^{\mathrm{t}}x\, {}^{\mathrm{t}}\!AAy = {}^{\mathrm{t}}(Ax)(Ay) = (Ax) \cdot (Ay) = T(x) \cdot T(y) = x \cdot y = {}^{\mathrm{t}}x\, y = {}^{\mathrm{t}}x\, Iy$$

for all x, $y \in \mathbb{R}^n$. But, as our next lemma show, this implies that ${}^{\mathrm{t}}\!AA = I$, as required.

Lemma. *Let F be any field and M, $N \in \operatorname{Mat}(n \times n, F)$. If ${}^{\mathrm{t}}x\, My = {}^{\mathrm{t}}x\, Ny$ for all x, $y \in F^n$, then $M = N$.*

Proof. We adopt the notation that if X is a matrix then X_{ij} denotes the (i, j)-entry of X. Similarly, if v is a column vector then v_j denotes the j-th entry of v.

Let e_1, e_2, …, e_n be the standard basis of F^n; that is,

$$e_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad e_2 = \begin{pmatrix} 0 \\ 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}, \quad \ldots, \quad e_n = \begin{pmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ 1 \end{pmatrix}.$$

For each $r \in \{1, 2, \dots, n\}$, the s-th entry of e_r is $(e_r)_s = \delta_{rs}$, the Kronecker delta. (Recall that δ_{rs} is 1 if $r = s$ and 0 if $r \neq s$). Now our hypothesis yields that ${}^{\mathrm{t}}e_i\, Me_j = {}^{\mathrm{t}}e_i\, Ne_j$ for all i, $j \in \{1, 2, \dots, n\}$; furthermore,

$$
{}^{\mathrm{t}}e_i\, Me_j = \sum_{k=1}^{n} \sum_{l=1}^{n} ({}^{t}e_i)_k M_{kl} (e_j)_l = \sum_{k=1}^{n} \sum_{l=1}^{n} \delta_{ik} M_{kl} \delta_{il} = M_{ij},
$$

since $\delta_{ik} M_{kl} \delta_{il}$ is zero unless $k = i$ and $l = j$, and similarly ${}^{\mathrm{t}}e_i\, Ne_j = N_{ij}$. Thus $M_{ij} = N_{ij}$ for all i and j, and so $M = N$. $\square$

A matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$ is said to be *orthogonal* if ${}^{\mathrm{t}}AA = I$. That is, an orthogonal matrix is a matrix whose transpose and inverse are equal.

Let $A \in \mathrm{Mat}(n \times n, \mathbb{R})$, and let a_1, a_2, …, a_n be the columns of A. That is,

$$
A = \left(a_1 \,\middle|\, a_2 \,\middle|\, \cdots \,\middle|\, a_n \right).
$$

Then the rows of ${}^{\mathrm{t}}A$ are ${}^{\mathrm{t}}a_1$, ${}^{\mathrm{t}}a_2, \dots,$ ${}^{\mathrm{t}}a_n$, and

$$
{}^{\mathrm{t}}AA = \begin{pmatrix} {}^{\mathrm{t}}a_1 \\ {}^{\mathrm{t}}a_2 \\ \vdots \\ {}^{\mathrm{t}}a_n \end{pmatrix} \left(a_1 \,\middle|\, a_2 \,\middle|\, \cdots \,\middle|\, a_n \right) = \begin{pmatrix} {}^{\mathrm{t}}a_1\, a_1 & {}^{\mathrm{t}}a_1\, a_2 & \dots & {}^{\mathrm{t}}a_1\, a_n \\ {}^{\mathrm{t}}a_2\, a_1 & {}^{\mathrm{t}}a_2\, a_2 & \dots & {}^{\mathrm{t}}a_2\, a_n \\ \vdots & \vdots & & \vdots \\ {}^{\mathrm{t}}a_n\, a_1 & {}^{\mathrm{t}}a_n\, a_2 & \dots & {}^{\mathrm{t}}a_n\, a_n \end{pmatrix}
$$

since the (i, j)-entry of ${}^{\mathrm{t}}AA$ is the product of the i-th row of ${}^{\mathrm{t}}A$ and the j-th column of A. Since ${}^{\mathrm{t}}a_i\, a_j = a_i \cdot a_j$, this can also be written as

$$
{}^{\mathrm{t}}AA = \begin{pmatrix} a_1 \cdot a_1 & a_1 \cdot a_2 & \dots & a_1 \cdot a_n \\ a_2 \cdot a_1 & a_2 \cdot a_2 & \dots & a_2 \cdot a_n \\ \vdots & \vdots & & \vdots \\ a_n \cdot a_1 & a_n \cdot a_2 & \dots & a_n \cdot a_n \end{pmatrix},
$$

and we conclude that ${}^{\mathrm{t}}AA = I$ if and only if $a_i \cdot a_j = \delta_{ij}$ for all i and j. That is, A is orthogonal if and only if the columns of A form an orthonormal basis of $\mathbb{R}^n$.

Summary

For any matrix $A \in \mathrm{Mat}(n \times n, \mathbb{R})$, the following conditions are equivalent:

- the columns of A form an orthonormal basis of $\mathbb{R}^n$;
- the map $T: \mathbb{R}^n \to \mathbb{R}^n$ defined by $x \mapsto Ax$ is an isometry;
- ${}^{\mathrm{t}}AA = I$.

Since ${}^{\mathrm{t}}AA = I$ if and and only if $A\,{}^{\mathrm{t}}A = I$ (given that A is square), another equivalent condition is that the columns of ${}^{\mathrm{t}}A$ form an orthonormal basis of $\mathbb{R}^n$ (or, if you prefer, the rows of A form an orthonormal basis of ${}^{\mathrm{t}}\mathbb{R}^n$). See Tutorial 7.

It is easy to generalize the above discussion to the complex case: it is simply necessary to take complex conjugates in a few places.

Definition. If A is a matrix over the complex field then we define the *conjugate* of A to be the matrix $\overline{A}$ whose (i,j)-entry is the complex conjugate of the (i,j)-entry of A (for all i and j).

An $n \times n$ complex matrix is said to be *unitary* if ${}^{\mathrm{t}}\overline{A}A = I$. Calculations similar to those above show that A is unitary if and only if $x \mapsto Ax$ is an isometry of $\mathbb{C}^n$, which in turn is equivalent to the columns of A forming an orthonormal basis of $\mathbb{C}^n$. See 5.15 of [VST].

A real valued function of n variables is said to be *analytic* at the origin O if its Taylor series at O converges to the function in some neighbourhood of O. Taking $n = 3$ for simplicity, we can then write

$$F(x,y,z) \approx a + (bx + cy + dz) + \tfrac{1}{2}(ex^2 + fy^2 + gz^2 + 2hxy + 2jxz + 2kyz) + \cdots$$

where the constant term a is $F(0,0,0)$, the coefficients b, c and d of the terms of degree 1 are given by the first order partial derivatives of F at the origin, and, in general, the coefficients of the degree n terms are given by the n-th order partial derivatives of F at the origin. The closer (x,y,z) is to $(0,0,0)$, the better the approximation. Furthermore, as (x,y,z) approaches $(0,0,0)$, the terms of degree two and higher become insignificant in comparison with the terms of degree one. So in a small enough neighbourhood of the origin

$$F(x,y,z) - F(0,0,0) \approx bx + cy + dz = L(x,y,z)$$

a linear function of (x,y,z). Thus we see that functions that are not linear may, nevertheless, be approximately linear. This is the basic reason for the importance of linear algebra.

A *stationary point* of the function F is a point where all its partial derivatives vanish. If the origin is a stationary point the linear function L above is identically zero, and so the terms of degree two assume a greater importance. We obtain

$$F(x,y,z) - F(0,0,0) \approx \tfrac{1}{2}Q(x,y,z)$$

where $Q(x,y,x)$ is a *homogeneous poynomial of degree 2*, or *quadratic form*:

$$Q(x,y,z) = ex^2 + fy^2 + gz^2 + 2hxy + 2jxz + 2kyz.$$

To understand the behaviour of functions in neighbourhoods of their stationary points it is necessary to understand the behaviour of of quadratic forms.

It is possible to apply matrix theory to the study of quadratic forms. The crucial observation is that

$$ex^2 + fy^2 + gz^2 + 2hxy + 2jxz + 2kyz = \begin{pmatrix} x & y & z \end{pmatrix} \begin{pmatrix} e & h & j \\ h & f & k \\ j & k & g \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

There is no need to restrict attention to three variables: the situation for any number of variables is totally analogous. If A is an $n \times n$ real symmetric matrix and x is the column vector whose n components are the variables $x_1, x_2, \ldots, x_n$, then

$$Q(x) = {}^{\mathrm{t}}xAx$$

is a quadratic form in $x_1, x_2, \ldots, x_n$. The coefficient of x_i^2 is A_{ii} and the coefficient of x_ix_j is $2A_{ij}$ (for all i and j).

In this context, the single most important fact is the following theorem.

Theorem. *If A is a real symmetric matrix then there exists an orthogonal matrix T such that ${}^{\mathrm{t}}TAT$ is diagonal.*

This is Theorem 5.19 (i) of [VST]. (The second part of Theorem 5.19, which gives the extension of this result to the complex case, was not covered in lectures.) The proof was omitted, but examples were given to show how one can go about finding an orthogonal matrix T to diagonalize a given symmetric matrix A. Furthermore, the key step in the proof *was* done in lectures. It goes as follows.

Let A be an $n \times n$ real symmetric matrix. Suppose that λ and μ are eigenvalues of A, with u and v corresponding eigenvectors. As we know from first year, λ and μ are roots of the characteristic polynomial of A. We shall show that in fact they must be real, but for the time being all we can say is that they are complex numbers that may or may not be real. Note that if an eigenvalue is a non-real complex number, the corresponding eigenvectors will also have non-real entries.

We have

$$\begin{aligned} Au &= \lambda u \\ Av &= \mu v \end{aligned}$$

where $\lambda, \mu \in \mathbb{C}$ and $u, v \in \mathbb{C}^n$, with $u \neq 0$ and $v \neq 0$. Now

$$\begin{aligned} \lambda(v \cdot u) = v \cdot (\lambda u) = v \cdot (Au) &= {}^{\mathrm{t}}\overline{v}Au \\ &= {}^{\mathrm{t}}\overline{v}\,{}^{\mathrm{t}}\overline{A}u \qquad \text{(since A is real and symmetric)} \\ &= ({}^{\mathrm{t}}\overline{Av})u = (Av) \cdot u = (\mu v) \cdot u = \overline{\mu}(v \cdot u), \end{aligned}$$

where we have made use of the fact that the dot product is linear in the second variable and conjugate linear in the first variable. Since the above holds, in particular, when $\mu = \lambda$ and $v = u$, it follows that

$$\lambda(v \cdot v) = \overline{\lambda}(v \cdot v).$$

Since $v \neq 0$, positive definiteness of the dot product yields that $v \cdot v \neq 0$. Hence we may cancel $v \cdot v$ from this equation, and deduce that $\lambda = \overline{\lambda}$. That is, λ is real. Since λ was an arbitrary eigenvalue of A, we have shown that all eigenvalues of a real symmetric matrix must be real.

Now suppose that λ and μ are distinct eigenvalues of the real symmetric matrix A, and u, v corresponding eigenvectors. In view of the fact that μ must be real, the equation we proved above becomes

$$\lambda(v \cdot u) = \mu(v \cdot u),$$

and since $\lambda \neq \mu$ (by hypothesis) we must have $v \cdot u = 0$. That is, v is orthogonal to u.

Summary

- Every eigenvalue of a real symmetric matrix is real.
- For a real symmetric matrix, eigenvectors belonging to distinct eigenvalues are orthogonal.

The remaining essential ingredient needed for the proof of Theorem 5.19 is the so-called "Fundamental Theorem of Algebra", which asserts that every non-constant polynomial with complex coefficients has a zero in $\mathbb{C}$. The proof of this fact is not part of the syllabus of this course, but (for interest only) a heuristic outline of a proof was presented in lectures. A separate document dealing with this has been prepared.

The Fundamental Theorem of Algebra guarantees that the characteristic polynomial of any matrix $A \in \mathrm{Mat}(n \times n, \mathbb{C})$ will have at least one root, and hence there exists at least one eigenvalue and a corresponding eigenvector. There exists an invertible matrix T such that $T^{-1}AT$ is diagonal if and only if it is possible to find n linearly independent eigenvectors. This can always be done it A is symmetric, but non-symmetric matrices are not necessarily diagonalizable. The following example shows illustrates this:

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

the characteristic polynomial of A is $(1-x)^3$, and so the only eigenvalue is 1. Any eigenvector will have to be a solution of

$$\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix},$$

but the general solution of this system involves only one arbitrary parameter: the solution space is 1-dimensional. So it is not possible to find a 3×3 matrix whose columns are linearly independent eigenvectors of A, and so A cannot be diagonalized.

For a symmetric matrix A it always turns out that the dimension of the λ-eigenspace equals the multiplicity of λ as a root of the characteristic equation.

Finding bases for the eigenspaces corresponding to all the eigenvalues will give us n vectors altogether, and the matrix T with these vectors as its columns will necessarily be invertible and will diagonalize A.

If we require T to be orthogonal then we need the columns of T to comprise an orthonormal basis of $\mathbb{R}^n$. After finding bases for the various eigenspaces we need to apply the Gram-Schmidt process to obtain orthonormal bases for these eigenspaces. Given that A is symmetric, it is automatically true that the distinct eigenspaces are orthogonal to each other, and so combining the orthonormal bases of the eigenspaces gives an orthonormal basis of $\mathbb{R}^n$.

Note that if A is not symmetric then eigenspaces corresponding to distinct eigenvalues are not orthogonal to each other. For example, let

$$A = \begin{pmatrix} 1 & 1 \\ 0 & 2 \end{pmatrix}.$$

The eigenvalues are 1 and 2. Both eigenspaces are one-dimensional, and are spanned, respectively, by the following vectors:

$$\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \qquad \begin{pmatrix} 1 \\ 1 \end{pmatrix}.$$

Observe that they are not orthogonal to each other.

Two examples were done in lectures to illustrate orthogonal diagonalization of symmetric matrices. Let

$$A = \begin{pmatrix} -2 & 1 & 1 \\ 1 & 6 & 3 \\ 1 & 3 & -4 \end{pmatrix}, \qquad B = \begin{pmatrix} 3 & 2 & 2 \\ 2 & 3 & 2 \\ 2 & 2 & 3 \end{pmatrix}.$$

We find that $\det(A - xI) = -(x+2)(x+5)(x-7)$ and $\det(B - xI) = (1-x)^2(7-x)$.

The three eigenspaces of A all have dimension 1, and they are spanned by the following three vectors (which have been normalized so as to have length 1):

$$\frac{1}{\sqrt{27}}\begin{pmatrix} 5 \\ -1 \\ 1 \end{pmatrix}, \qquad \frac{1}{\sqrt{18}}\begin{pmatrix} 1 \\ 1 \\ -4 \end{pmatrix}, \qquad \frac{1}{\sqrt{54}}\begin{pmatrix} 1 \\ 7 \\ 2 \end{pmatrix}.$$

The 1-eigenspace of B has dimension 2, and consists of all vectors whose entries add to zero. We find that

$$\frac{1}{\sqrt{3}}\begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}, \qquad \frac{1}{\sqrt{2}}\begin{pmatrix} -1 \\ 1 \\ 0 \end{pmatrix}, \qquad \frac{1}{\sqrt{6}}\begin{pmatrix} 1 \\ -2 \\ 1 \end{pmatrix}$$

is an orthonormal basis of $\mathbb{R}^3$ consisting of a 7-eigenvector and two 1-eigenvectors of B. The two matrices with these orthonormal bases for their columns are orthogonal and diagonalize A and B respectively.

Let $U \in \mathrm{Mat}(n \times n, \mathbb{C})$ be a unitary matrix. Then the map $\mathbb{C}^n \to \mathbb{C}^n$ given by multiplication by U is an isometry; so

$$\|Uv\| = \|v\| \qquad \text{for all } v \in \mathbb{C}^n.$$

If v is an eigenvector of U corresponding to the eigenvalue λ, then $v \neq 0$ and $Uv = \lambda v$. The above equation becomes

$$\|\lambda v\| = \|v\|,$$

and since we know in general that $\|\lambda v\| = |\lambda|\,\|v\|$, and $\|v\| \neq 0$ since $v \neq 0$, it follows that $|\lambda| = 1$. We have shown that every (complex) eigenvalue λ of a unitary matrix lies on the unit circle (the set of numbers in the complex plane of modulus 1).

An orthogonal matrix is exactly the same thing as a unitary matrix that happens to have real entries; so it is also true that all complex eigenvalues of a real orthogonal matrix must lie on the unit circle. In particular, the only possible real eigenvalues are ± 1.

Now let P be a 3×3 orthogonal matrix. The characteristic polynomial of P is

$$\det(P - xI) = (\lambda_1 - x)(\lambda_2 - x)(\lambda_3 - x)$$

where λ_1, λ_2, λ_3 are the eigenvalues of P. Note that putting $x = 0$ gives $\det P = \lambda_1\lambda_2\lambda_3$. Since a real cubic polynomial must have at least one real root, P has at least one real eigenvalue. So in view of the argument above we may assume that $\lambda_1 = \pm 1$. If λ_2 and λ_3 are not real then we must have $\lambda_2 = \alpha$ and $\lambda_3 = \overline{\alpha}$, where α lies on the unit circle. Otherwise we must have $\lambda_2 = \pm 1$ and $\lambda_3 = \pm 1$ also. In the former case we find that

$$\det P = \lambda_1\lambda_2\lambda_3 = \lambda_1|\alpha|^2 = \lambda_1,$$

while in the latter case $\det P = 1$ if the multiplicity of 1 as an eigenvalue is 1 or 3, and $\det P = -1$ if the multiplicity of 1 as an eigenvalue is 0 or 2. In either case we see that $\det P = \pm 1$, and if $\det P = 1$ then 1 is an eigenvalue of P.

Our aim is to give a geometrical description of the transformation $\mathbb{R}^3 \to \mathbb{R}^3$ given by multiplication by P. We may as well assume that $\det P = 1$, since $\det(-P) = \det(-I)\det P = (-1)^3 \det P = -\det P$, so that replacing P by $-P$ changes the sign of the determinant. From our discussion above, we know that 1 is an eigenvalue; so we may choose a vector u such that $Pu = u$. The set $\mathcal{L}$ consisting of all scalar multiples of u is a line through the origin in $\mathbb{R}^3$. We define

$$\mathcal{P} = \{\, v \in \mathbb{R}^3 \mid v \cdot u = 0 \,\},$$

the plane through the origin perpendicular $\mathcal{L}$.

If $v \in \mathcal{P}$ then since multiplication by P preserves the dot product, we see that

$$(Pv) \cdot u = (Pv) \cdot (Pu) = v \cdot u$$

(in view of the fact that $Pu = u$). So $Pv \in \mathcal{P}$ also, and we conclude that multiplication by P gives rise to a transformation of the plane $\mathcal{P}$. Moreover, this transformation preserves lengths (since $x \mapsto Px$ is an isometry of $\mathbb{R}^3$). It is, of course, also a linear transformation. Now the only length preserving linear transformations of a plane are reflections and rotations. In this case the transformation of $\mathcal{P}$ must actually be a rotation rather than a reflection: if it were a reflection then we could find an eigenvector of P lying in the plane $\mathcal{P}$ and belonging to the eigenvalue -1, and another belonging to the eigenvalue 1. This would mean that $\lambda_1\lambda_2\lambda_3 = -1$, contrary to our assumption that $\det P = 1$.

We conclude that the transformation $T: x \mapsto Px$ fixes all vectors on the line $\mathcal{L}$ and acts as a rotation on $\mathcal{P}$, the plane through the origin perpendicular to $\mathcal{L}$. Since any vector in $\mathbb{R}^3$ can be expressed as the sum of a vector in $\mathcal{P}$ and a vector in $\mathcal{L}$, we can deduce from this that T is a rotation of $\mathbb{R}^3$ about the axis $\mathcal{L}$.

As an application of this idea, consider a surface in $\mathbb{R}^3$ with equation of the form

$$ex^2 + fy^2 + gz^2 + 2hxy + 2jxz + 2kyz = C.$$

As we have seen, we can write this as ${}^{\mathrm{t}}\underset{\sim}{x} A \underset{\sim}{x} = C$, where A is a symmetric matrix, and we can choose an orthogonal matrix P that diagonalizes the matrix. Moreover, if P diagonalizes A then so does $-P$; so we can assume that $\det P = 1$. Now the change of variables

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = P \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}$$

corresponds to a rotation of the coordinate axes; moreover,

$$\begin{pmatrix} x & y & z \end{pmatrix} A \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} X & Y & Z \end{pmatrix} {}^{\mathrm{t}}PAP \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} = \lambda_1 X^2 + \lambda_2 Y^2 + \lambda_3 Z^2,$$

where λ_1, λ_2, λ_3 are the eigenvalues of A. It is now relatively easy to determine what the surface is like. See Tutorial 8 for more details.