MATH 340: EIGENVECTORS, SYMMETRIC MATRICES, AND ORTHOGONALIZATION

Let A be an $n \times n$ real matrix. Recall some basic definitions.

• A is *symmetric* if $A^t = A$;
• A vector $\mathbf{x} \in \mathbb{R}^n$ is an *eigenvector* for A if $\mathbf{x} \neq \mathbf{0}$, and if there exists a number λ such that $A\mathbf{x} = \lambda\mathbf{x}$. We call λ the *eigenvalue* corresponding to $\mathbf{x}$;
• We say a set of vectors $\mathbf{v}_1, \ldots, \mathbf{v}_k$ in $\mathbb{R}^n$ is *orthogonal* if $\mathbf{v}_i \cdot \mathbf{v}_j = 0$ whenever $i \neq j$. We say the vectors are *orthonormal* if in addition each $\mathbf{v}_i$ is a unit vector.
• We say P is *orthogonal* if $P^tP = I$ (Thus, P is invertible, and $P^{-1} = P^t$). Equivalently, the columns (or rows) of P form an orthonormal set.

We want to prove (in a not-too-painful way!) the following very important theorem.

Theorem 1 (Principal axis theorem)**.** *The following statements are equivalent:*
 (i) *A is symmetric;*
 (ii) *There exists an orthonormal basis for $\mathbb{R}^n$ consisting of eigenvectors of A;*
 (iii) *There exists an orthogonal matrix P such that P^tAP is diagonal.*

We can prove some parts of the theorem right away without much work.

$(iii) \Rightarrow (i)$: Assume P exists as in (iii), and write $P^tAP = D$, where D is diaganal. Note that $P^{-1} = P^t$ implies that $A = PDP^t$. Now A is symmetric follows from

$$A^t = (PDP^t)^t = P^{tt}D^tP^t = PDP^t = A.$$

(We used D diagonal to justify $D^t = D$ here).

$(ii) \Rightarrow (iii)$: Suppose $\mathbf{v}_1, \cdots, \mathbf{v}_n$ are an orthonormal basis of eigenvectors for A. Let P be the matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$; in other words $P\mathbf{e}_i = \mathbf{v}_i$ for each i.

Claim: P is orthogonal.
Pf: Let δ_{ij} denote the Kronecker symbol: $\delta_{ij} = 0$ if $i \neq j$ and $\delta_{ii} = 1$. We have

$$\text{the } \mathbf{v}_i\text{'s are orthonormal} \Leftrightarrow \mathbf{v}_i \cdot \mathbf{v}_j = \delta_{ij}\ (\forall i, j) \Leftrightarrow \mathbf{e}_i^t P^t P \mathbf{e}_j = \delta_{ij}\ (\forall i, j) \Leftrightarrow P^tP = I,$$

which shows that P is orthogonal, proving the claim.

Next, we claim that P^tAP is diagonal. For each i, we have $A\mathbf{v}_i = \lambda_i\mathbf{v}_i$ for some scalar λ_i. Using $\mathbf{v}_i = P\mathbf{e}_i$ and $P^t = P^{-1}$, this gives

$$AP\mathbf{e}_i = \lambda_i P\mathbf{e}_i$$

and thus

$$P^tAP\mathbf{e}_i = \lambda_i\mathbf{e}_i.$$

This is the same as saying that $P^tAP = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, a diagonal matrix with the λ_i's down the diagonal. This proves the implication $(ii) \Rightarrow (iii)$.

(iii) $\Rightarrow$ *(ii)*: This is similar to the above implication. Assume P exists as in (iii), and define $\mathbf{v}_i = P\mathbf{e}_i$. Then the assumption that P is orthogonal implies that the $\mathbf{v}_i$'s form an orthonormal set. Further, the identity

$$P^t A P \mathbf{e}_i = \lambda_i \mathbf{e}_i,$$

which follows from $P^tAP = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$, together with $P^{-1} = P^t$, shows that $A\mathbf{v}_i = \lambda_i \mathbf{v}_i$, for each i. Thus, we have found an orthonormal basis of eigenvectors for A.

It remains to prove *(i)* $\Rightarrow$ *(iii)*. This is the hardest and most interesting part. I will proceed here in a different manner from what I explained (only partially) in class.

The main ingredient is the following proposition.

Proposition 2. *Any symmetric matrix A has an eigenvector.*

Remark: In the end, we will see that in fact A will have a lot more than just one eigenvector, but since the proof of *(i)* $\Rightarrow$ *(iii)* is ultimately done by a kind of induction, we need to produce a first eigenvector to "get started". It is not at all the case that an arbitrary matrix has an eigenvector. For example, suppose A is the matrix

$$R_\theta = \begin{bmatrix} \cos(\theta) & -\sin(\theta) \\ \sin(\theta) & \cos(\theta) \end{bmatrix},$$

the matrix corresponding to "rotation by θ radians". If $\theta \neq 0$, it is geometrically clear that there is no non-zero vector $\mathbf{x}$ such that $\mathbf{x}$ and $R_\theta(\mathbf{x})$ are colinear; thus, R_θ has no eigenvectors at all.

There are several approaches to this proposition, each important in its own right. We will settle for the one which is closest to the material we have already covered in this class.

Our Method: the Optimization method (cf. Hubbard/Hubbard, page 372-373).
Let $Q_A(\mathbf{x}) = \mathbf{x}^t A \mathbf{x}$, the quadratic form corresponding to A. Explicitly, if $A = (a_{ij})$, then $Q_A(x_1, \ldots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j$. It is obviously a real-valued differentiable function of n variables. One can show its derivative at $\mathbf{a} = (a_1, \ldots, a_n)^t$ is the linear transformation $\mathbf{h} \mapsto \mathbf{a}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{a} = 2\mathbf{a}^t A \mathbf{h}$ (the last equality since $A^t = A$), so that

$$DQ_A(\mathbf{a})\mathbf{h} = \mathbf{a}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{a} = 2\mathbf{a}^t A \mathbf{h}.$$

Indeed, we just need to note that the following holds

$$\lim_{\mathbf{h}\to\mathbf{0}} \frac{(\mathbf{a}+\mathbf{h})^t A(\mathbf{a}+\mathbf{h}) - \mathbf{a}^t A \mathbf{a} - (\mathbf{a}^t A \mathbf{h} + \mathbf{h}^t A \mathbf{a})}{|\mathbf{h}|} = 0$$

(you should think this over carefully). Another way to phrase this result is

$$\nabla Q_A(\mathbf{a}) = 2\mathbf{a}^t A.$$

Now we consider the unit sphere S in $\mathbb{R}^n$: the unit sphere consists of vectors of length 1, i.e.,

$$S = \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x}| = 1\}.$$

This set is closed and bounded. We are going to try to maximize $Q_A(\mathbf{x})$ subject to the constraint that $\mathbf{x} \in S$, ie., subject to $g(\mathbf{x}) = 1$, where

$$g(\mathbf{x}) = |\mathbf{x}|^2 = x_1^2 + \cdots + x_n^2.$$

We have (think of $\mathbf{x}$ as a column vector, so $\mathbf{x}^t$ is the row vector $(x_1, \ldots, x_n)$)

$$\nabla g(\mathbf{x}) = 2\mathbf{x}^t.$$

By general theorems (we skipped these – see sec. 1.6 of your text), any continuous function such as Q_A attains an absolute maximum on a closed and bounded set such as S. Let $\mathbf{v}$ be a point in the sphere S where Q_A attains its maximum value on the sphere S. By the theory of Lagrange multipliers, there is some scalar $\lambda \in \mathbb{R}$ such that

$$\nabla Q_A(\mathbf{v}) = \lambda \nabla g(\mathbf{v}),$$

i.e.,

$$2\mathbf{v}^t A = \lambda 2\mathbf{v}^t,$$

which, since $A^t = A$, yields the following on applying transpose to both sides:

$$A\mathbf{v} = \lambda \mathbf{v}.$$

This shows that $\mathbf{v}$ is indeed an eigenvector for A. This proves the proposition.

Here is a sketch of another important method for proving the proposition.

Another method: find a root of the characteristic polynomial for A.

Here, the *characteristic polynomial* for A is the polynomial $\det(A - tI)$, a real polynomial of degree n in the variable t. We can regard A as a matrix over the complex numbers $\mathbb{C}$, and also we can regard, in the usual way, the matrix A as a linear transformation $A : \mathbb{C}^n \to \mathbb{C}^n$. The definitions of eigenvalue and eigenvector make sense for complex matrices – the definitions are the same as before.

Lemma 0.1. *Let A be an $n \times n$ matrix over $\mathbb{C}$. Then:*

(a) *$\lambda \in \mathbb{C}$ is an eigenvalue corresponding to an eigenvector $\mathbf{x} \in \mathbb{C}^n$ if and only if λ is a root of the characteristic polynomial $\det(A - tI)$;*
(b) *Every complex matrix has at least one complex eigenvector;*
(c) *If A is a real symmetric matrix, then all of its eigenvalues are real, and it has a real eigenvector (ie. one in the subset $\mathbb{R}^n \subset \mathbb{C}^n$).*

The third part of this Lemma gives us another proof of the Proposition above.

Proof. (a) Fix a complex number λ. There is a non-zero complex vector $\mathbf{x}$ such that $A\mathbf{x} = \lambda\mathbf{x}$ if and only if $A - \lambda I$ has non-zero null space, which holds if and only if the determiniant $\det(A - \lambda I)$ vanishes. This gives the desired equivalence.

(b) The fundamental theorem of algebra asserts that any complex polynomial has a root in the complex numbers. Applying this to the polynomial $\det(A - tI)$, we see it has a root in the complex numbers. By invoking part (a), we see A possesses an eigenvalue, that is a number λ such that there is some non-zero complex vector $\mathbf{x}$ with $A\mathbf{x} = \lambda\mathbf{x}$.

(c) First of all, by part (b), we know A has at least a complex eigenvalue. Once we show this is necessarily real, then the same argument as in the part (a) shows that A has a real eigenvector, and we'll have proved the proposition another way.

It remains to show that if $a+ib$ is a complex eigenvalue for the real symmetric matrix A, then $b = 0$, so the eigenvalue is in fact a real number. Suppose $\mathbf{v} + i\mathbf{w} \in \mathbb{C}^n$ is a complex eigenvector with eigenvalue $a + ib$ (here $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$). Note that applying the complex conjugation to the identity $A(\mathbf{v} + i\mathbf{w}) = (a + ib)(\mathbf{v} + i\mathbf{w})$ yields $A(\mathbf{v} - i\mathbf{w}) = (a - ib)(\mathbf{v} - i\mathbf{w})$. We will show $b = 0$ by considering

$$(\mathbf{v} - i\mathbf{w})^t A(\mathbf{v} + i\mathbf{w}).$$

On the one hand, this is

$$(\mathbf{v} - i\mathbf{w})^t(A(\mathbf{v} + i\mathbf{w})) = (a + ib)((\mathbf{v} - i\mathbf{w})^t(\mathbf{v} + i\mathbf{w})) = (a + ib)(|\mathbf{v}|^2 + |\mathbf{w}|^2).$$

On the other hand, a similar calculation show it is

$$((\mathbf{v} - i\mathbf{w})^t A)(\mathbf{v} + i\mathbf{w}) = (A(\mathbf{v} - i\mathbf{w}))^t(\mathbf{v} + i\mathbf{w})) = (a - ib)(|\mathbf{v}|^2 + |\mathbf{w}|^2).$$

Putting these together and noting that $|\mathbf{v}|^2 + |\mathbf{w}|^2 \neq 0$, we get $a + ib = a - ib$, hence $b = 0$. This completes the proof of part (c), and thus the Lemma.

□

Finally, we return to the implication $(i) \Rightarrow (iii)$, having proved the Proposition now in two different ways:

Proof. Suppose A is symmetric. We want to show there is an orthonormal matrix P such that P^tAP is diagonal. According to the proposition, there is an eigenvector $\mathbf{u}_1$ with eigenvalue λ_1. We may as well assume $\mathbf{u}_1$ is a unit vector (if necessary, divide it by its length).

Let us proceed by induction on n, the size of A. We prove the "bootstrap step": we assume that every symmetric matrix of size $n - 1$ has an orthogonal Q which diagonalizes it, and we will deduce that A may also be diagonalized by an orthogonal matrix.

Let us consider $\mathbf{u}_1$ as the first element in a basis $\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_n$ of $\mathbb{R}^n$ (it is always possible to find the other basis elements $\mathbf{u}_2, \ldots, \mathbf{u}_n$). Now apply the Gram-Schmidt orthogonalization process to this basis to produce an orthonormal basis $\mathbf{v}_1 = \mathbf{u}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n$. Let P_1 be the orthogonal matrix whose columns are $\mathbf{v}_1, \ldots, \mathbf{v}_n$.

Note that the first column of $P_1^{-1}AP_1 = P_1^tAP_1$ is $[\lambda_1, 0, \ldots, 0]^t$ (Here's why: $P_1^{-1}AP_1\mathbf{e}_1 = P_1^{-1}A\mathbf{u}_1 = P_1^{-1}\lambda_1\mathbf{u}_1 = \lambda_1\mathbf{e}_1$.) Since $P_1^tAP_1$ is symmetric (why?), this means it can be written

$$\begin{bmatrix} \lambda_1 & 0 \\ 0 & B \end{bmatrix}$$

where B is an $n - 1 \times n - 1$ symmetric matrix, and the 0 denotes a string of $n - 1$ zeros in the first row (resp. first column).

By our induction hypothesis, there exists an orthogonal matrix Q such that Q^tBQ is diagonal. Then we easily see that if we set

$$P = P_1 \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix},$$

then P is orthogonal and P^tAP is diagonal. This completes the proof of $(i) \Rightarrow (iii)$.

$\square$