A proof of the spectral theorem for symmetric matrices (Optional)

Math 419

In class we have covered - and by now seen some applications of - the following result

Theorem 1 (The spectral theorem – for symmetric matrices). *If $A : \mathbb{R}^n \to \mathbb{R}^n$ is a symmetric matrix, then*

- *A has eigenvectors $v_1, ..., v_n$ such that $\{v_1, ..., v_n\}$ is an orthonormal basis for $\mathbb{R}^n$.*
- *A has real eigenvalues $\lambda_1, ..., \lambda_n$.*

In fact the first bullet implies the second (think about why), and moreover these bullets have the implication that

$$A = Q\Lambda Q^{-1},$$

where

$$Q = \begin{pmatrix} | & & | \\ v_1 & \cdots & v_n \\ | & & | \end{pmatrix} \text{ is an } \textit{orthogonal matrix}, \text{ and } \quad \Lambda = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

We have by now seen this theorem applied to positive definite matrices and a first glance of it being applied to the singular value decomposition. These are important ideas!

Nonetheless in class we gave a proof of this theorem only when A has all distinct eigenvalues, in which case we already know that A is diagonalizable. (If you are in the first section you suffered through a very rushed outline of a proof of the spectral theorem in general that I think was probably incomprehensible.)

In this note I want to give a proof of Theorem 1 in general, even in the case of repeated eigenvalues. The best proof requires a little bit more abstraction that we usually make use of in class.

Reading through the proof below is completely optional for Math 419. If we haven't covered them otherwise, the ideas I make use of here won't appear on any exam. Nonetheless it is always nice to know why something that has been claimed is true!

Recall that we showed in class

Lemma 1. *If $A : \mathbb{R}^n \to \mathbb{R}^n$ is symmetric if and only if*

$$x \cdot Ay = y \cdot Ax, \quad \textit{for all}\, x, y \in \mathbb{R}^n.$$

This motivates the following some more abstract definition.

Definition 1. *Let W be a subspace of $\mathbb{R}^N$. For a linear transformation $A : W \to W$, we say that A is symmetric (in the subspace W) if*

$$x \cdot Ay = y \cdot Ax \quad \textit{for all}\, x, y \in W.$$

(Note that $A : W \to W$ just means that for any $w \in W$, we also have $Aw \in W$.)

We also need the following lemma (you have used/reproved a variant of it Thursday in class in discussing positive definite matrices and the singular value decomposition):

Lemma 2. *If W is a n-dimensional subspace of $\mathbb{R}^N$ and $u_1, ..., u_n$ is an orthonormal basis for W, then for*

$$x = \alpha_1 u_1 + \cdots + \alpha_n u_n$$
$$y = \beta_1 u_1 + \cdots + \beta_n u_n,$$

we have

$$x \cdot y = \alpha_1 \beta_1 + \cdots + \alpha_n \beta_n.$$

Proof. Note

$$(\alpha_1 u_1 + \cdots \alpha_n u_n) \cdot (\beta_1 u_n + \cdots + \beta_n u_n) = \sum_{i,j} \alpha_i \beta_j (u_i \cdot u_j) = \sum_i \alpha_i \beta_i,$$

as $u_i \cdot u_j = 0$ for $i \neq j$ and $u_i \cdot u_i = |u_i|^2 = 1$. □

Our proof of the spectral theorem now involves some calculus; we need to recall i) the product rule for differentiating dot products, and ii) Lagrange multipliers.

Lemma 3 (Product rule). *Let $a(t)$ and $b(t)$ be $\mathbb{R}^n$ valued functions, (that is $a : \mathbb{R} \to \mathbb{R}^n$ and $b : \mathbb{R} \to \mathbb{R}^n$) with each coordinate differentiable. Then*

$$\frac{d}{dt}\big(a(t) \cdot b(t)\big) = a'(t) \cdot b(t) + a(t) \cdot b'(t).$$

Lemma 4 (Lagrange multipliers). *Suppose $f : \mathbb{R}^n \to \mathbb{R}$ and $g : \mathbb{R}^n \to \mathbb{R}$, both having continuous partial derivatives in all variables. Consider the problem*

$$\text{Maximize } f(x_1, ..., x_n)$$
$$\text{Subject to } g(x_1, ..., x_n) = 0.$$

If x^* is the point at which a maximum occurs, then for all i,

$$\partial_i f(x_1^*, ..., x_n^*) = \lambda \partial_i g(x_1^*, ..., x_n^*),$$

for some $\lambda \in \mathbb{R}$.

(We do not prove these lemma here, but a proof can be found in good books on vector calculus.)

We can now turn to the more abstract version of the spectral theorem.

THEOREM 2 (An abstract spectral theorem). *Let $n \leq N$ and let W be an n-dimensional subspace of $\mathbb{R}^N$. For a linear transformation $A : W \to W$ that is symmetric (in the subspace W), we have*

- *There are eigenvectors $v_1, ..., v_n \in W$ of A such that $\{v_1, ..., v_n\}$ is an orthonormal basis for W.*
- *Associated to each eigenvector v_i is a real eigenvalue λ_i.*

Note that this implies Theorem 1 by letting $n = N$ and $W = \mathbb{R}^n$.

PROOF OF THEOREM 2. Our proof is by induction on n. (That is prove the result for $n = 1$ and all N, and then note that if the result is true for some $n-1$ and all N, it is also true for n and all N.) For $n = 1$, check that the result is obvious. Suppose now that we have proved the result for $n-1$; we will show it's also true for n.

For $n > 1$, we let $\{u_1, ..., u_n\}$ be an arbitrary orthonormal basis for W. (In particular, our choice of $\{u_1, ..., u_n\}$ has nothing to do with the transformation A.) Represent an arbitrary element of W by

$$w = \alpha_1 u_1 + \cdots + \alpha_n u_n. \tag{1}$$

We think of w as being a function of the 'coordinates' $\alpha_1, ..., \alpha_n$ and indeed we could write $w = w(\alpha_1, ..., \alpha_n)$. We consider the following problem:

$$\text{Maximize } w \cdot Aw$$

$$\text{Subject to } w \cdot w = 1 \quad (\text{equivalently } w \cdot w - 1 = 0)$$

Let

$$F(\alpha_1, ..., \alpha_n) = w \cdot Aw$$

$$G(\alpha_1, ..., \alpha_n) = w \cdot w - 1.$$

The method of Lagrange multipliers implies at the maximum (α_i^*), for each i,

$$\partial_i F(\alpha_1^*, ..., \alpha_n^*) = \lambda \partial_i G(\alpha_1^*, ..., \alpha_n^*). \tag{2}$$

But,

$$\begin{aligned}\partial_i F(\alpha_1, ..., \alpha_n) = \frac{\partial}{\partial \alpha_i} w \cdot Aw &= (\frac{\partial}{\partial \alpha_i} w) \cdot Aw + w \cdot (\frac{\partial}{\partial \alpha_i} Aw) \quad \text{(product rule)} \\ &= (\frac{\partial}{\partial \alpha_i} w) \cdot Aw + w \cdot (A \frac{\partial}{\partial \alpha_i} w) \quad (A \text{ is linear}) \\ &= u_i \cdot Aw + w \cdot Au_i \quad \text{(compute based on (1))} \\ &= 2u_i \cdot Aw.\end{aligned}$$

By a similar process,

$$\partial_i G(\alpha_1, ..., \alpha_n) = \frac{\partial}{\partial \alpha_i}(w \cdot w - 1) = 2u_i \cdot w.$$

Hence if

$$w^* = \alpha_1^* u_1 + \cdots + \alpha_n^* u_n$$

is where the maximum occurs for our optimization problem, then (2) implies

$$u_i \cdot Aw^* = \lambda u_i \cdot w^*. \tag{3}$$

If we have

$$Aw^* = \beta_1^* u_1 + \cdots + \beta_n^* u_n,$$

From Lemma 2, (3) implies

$$\beta_i^* = \lambda \alpha_i^* \quad \text{for all } i,$$

and hence

$$Aw = \lambda w.$$

That is to say, at the value w^* that maximizes our optimization problem, we must have that w^* is an eigenvector of A, with (obviously real) eigenvalue λ.

For w^* achieving the maximum, let $u_1 = w^*$ and $\lambda_1 = \lambda$. Then consider the $(n-1)$-dimensional subspace

$$W' = \text{span}(u_1)^{\perp}.$$

I claim that $A : W' \to W'$; that is for any $w' \in W'$, we have $A(w') \in W'$. This is the case for the following reason: if

$$0 = w' \cdot u_1,$$

then

$$0 = w' \cdot \lambda u_1 = w' \cdot Au_1 = Aw' \cdot u_1,$$

and this establishes $A(w') \in W'$.

By induction, the $(n-1)$-dimensional space W' has an orthonormal basis $\{u_2, ..., u_n\}$ of eigenvectors of A, each with a real eigenvalue, and u_1 is a unit vector orthogonal to each of $u_2, ..., u_n$, so $\{u_1, ..., u_n\}$ is an orthonormal basis of eigenvectors of W, and this proves the claim. □