**Symmetric Matrices and the Spectral Theorem**

(I) An $n \times n$ square matrix is said to be **symmetric** if its entries are symmetric across the main diagonal. For $n = 2$ and $n = 3$ these are the matrices of the form

$$\begin{bmatrix} a & b \\ b & c \end{bmatrix}, \qquad \begin{bmatrix} a & b & c \\ b & d & e \\ c & e & f \end{bmatrix}.$$

(II) If $A$ is symmetric, then

$$\vec{v} \cdot (A\vec{w}) = (A\vec{v}) \cdot \vec{w} \tag{1}$$

for all vectors $\vec{v}$ and $\vec{w}$. Indeed, being symmetric is equivalent to the condition that $A = A^T$, which allows us to write

$$\vec{v} \cdot (A\vec{w}) = \vec{v}^T A\vec{w} = \vec{v}^T A^T \vec{w} = (A\vec{v})^T \vec{w} = (A\vec{v}) \cdot \vec{w}.$$

(III) The above equation yields something interesting if $\vec{v}$ and $\vec{w}$ are both eigenvectors. If

$$A\vec{v} = \lambda\vec{v}, \qquad \text{and} \qquad A\vec{w} = \mu\vec{w},$$

then substituting into (1) yields

$$\mu\vec{v} \cdot \vec{w} = \lambda\vec{v} \cdot \vec{w},$$

or

$$(\mu - \lambda)\vec{v} \cdot \vec{w} = 0.$$

Hence, if $\mu \neq \lambda$, then $\vec{v}$ is orthogonal to $\vec{w}$.

(IV) The most important deeper property of symmetric matrices is that a symmetric matrix with real entries is always diagonalizable and all the eigenvalues and eigenvectors are real. This is in contrast to more complicated cases like

$$\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \qquad \text{and} \qquad \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix},$$

the first of which is defective, having only the eigenvectors $\begin{bmatrix} x_1 \\ 0 \end{bmatrix}$, and the second of which has complex eigenvalues and eigenvectors.

(V) One version of the Spectral Theorem says that given any $n \times n$ real symmetric matrix $A$, there is an orthonormal basis of $\mathbb{R}^n$ consisting of eigenvectors of $A$. We prove this using multivariable calculus. The first step is maximizing

$$(A\vec{v}) \cdot \vec{v},$$

subject to the constraint

$$\|\vec{v}\|^2 = 1.$$

Kiril Datchev, May 4, 2025. Questions, comments, and corrections are gratefully received at kdatchev@purdue.edu.

By the method of Lagrange multipliers, [1] this maximum occurs at a point where the gradient of the function being maximized is a constant multiple of the gradient of the constraint, i.e.

$$\nabla(A\vec{v})\cdot\vec{v} = \lambda\nabla(\|\vec{v}\|^2 - 1). \tag{4}$$

To compute the gradients, we expand:

$$\|\vec{v}+\vec{u}\|^2 = \|\vec{v}\|^2 + 2\vec{v}\cdot\vec{u} + \|\vec{u}\|^2,$$

and recall the Taylor expansion formula

$$F(\vec{v}+\vec{u}) = F(\vec{v}) + \nabla F(\vec{v})\cdot\vec{u} + \cdots,$$

which yields

$$\nabla\|\vec{v}\|^2 = 2\vec{v}. \tag{5}$$

Similarly,

$$(A(\vec{v}+\vec{u}))\cdot(\vec{v}+\vec{u}) = (A\vec{v})\cdot\vec{v} + 2(A\vec{v})\cdot\vec{u} + 2(A\vec{u})\cdot\vec{u},$$

where we used (1) to simplify, and hence

$$\nabla(A\vec{v})\cdot\vec{v} = 2A\vec{v}. \tag{6}$$

Substituting (5) and (6) into (4) yields

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

Thus the solution to the maximization problem is an eigenvector, and the Lagrange multiplier is an eigenvalue.

---

[1] See Paul's Notes for an introduction to Lagrange multipliers with pictures and examples. A neat derivation of the general formula comes from linear algebra. Specifically, if $\vec{v}$ is a maximizer of $F(\vec{v})$ subject to the constraint that it lies in the surface $M$ given by the equations $G_1(\vec{v}) = \cdots = G_k(\vec{v}) = 0$, then the Lagrange multipliers formula says that there are constants $\lambda_1, \ldots, \lambda_k$ such that

$$\nabla F(\vec{v}) = \lambda_1\nabla G_1(\vec{v}) + \cdots + \lambda_k\nabla G_k(\vec{v}). \tag{2}$$

To derive this, we start with the more general formula

$$\nabla F(\vec{v}) = \lambda_1\nabla G_1(\vec{v}) + \cdots + \lambda_k\nabla G_k(\vec{v}) + \vec{w}, \tag{3}$$

where $\vec{w}$ is tangent to the surface $M$; this comes from the fact that *any* vector is a sum of a term perpendicular to the surface (denoted here by the general linear combination $\lambda_1\nabla G_1(\vec{v}) + \cdots + \lambda_k\nabla G_k(\vec{v})$) a term tangent to the surface (denoted here by $\vec{w}$).

Next, let $\vec{\gamma}(t)$ be a curve in the surface $M$ such that $\vec{\gamma}(0) = \vec{v}$ and $\vec{\gamma}'(0) = \vec{w}$. Since $\vec{v}$ is a maximizer of $F$, it follows that 0 is a maximizer of $h(t) = F(\vec{\gamma}(t))$. Consequently

$$0 = h'(0) = \nabla F(\vec{\gamma}(0))\cdot\vec{\gamma}(0) = \nabla F(\vec{v})\cdot\vec{w} = \|\vec{w}\|^2,$$

where for the last equality we substituted (3) and used the fact that $\nabla G_j(\vec{v})\cdot\vec{w} = 0$ for every $j$. Hence $\vec{w} = 0$, which reduces (3) to (2).

A nice way to encapsulate (2) is to say that constrained maxima occur at critical points of the corresponding *Lagrangian function*, defined by

$$\mathcal{F}(\vec{v}, \lambda_1, \ldots, \lambda_k) = F(\vec{v}) - \lambda_1 G_1(\vec{v}) - \cdots - \lambda_k G_k(\vec{v}).$$

We will see that what we have found is the *largest* eigenvalue of $A$. [2] To find the next largest, we do the same maximization subject to an additional constraint: we label the above eigenvector and eigenvalue as $\vec{v}_1$ and $\lambda_1$, and we wish to maximize

$$(A\vec{v}_2) \cdot \vec{v}_2,$$

subject to the constraints

$$\|\vec{v}_2\|^2 = 1, \qquad \vec{v}_2 \cdot \vec{v}_1 = 0.$$

This time the method of Lagrange multipliers says that at a maximum we have

$$\nabla(A\vec{v}_2) \cdot \vec{v}_2 = \lambda \nabla(\|\vec{v}_2\|^2 - 1) + \mu \nabla(\vec{v}_2 \cdot \vec{v}_1),$$

(with gradients taken with respect to $\vec{v}_2$) or

$$A\vec{v}_2 = \lambda \vec{v}_2 + \mu \vec{v}_1. \tag{7}$$

We compute $\mu$ by dotting both sides with $\vec{v}_1$, to get

$$(A\vec{v}) \cdot \vec{v}_1 = \lambda \vec{v} \cdot \vec{v}_1 + \mu \vec{v}_1 \cdot \vec{v}_1 = \mu \vec{v}_1 \cdot \vec{v}_1, \tag{8}$$

where for the second equals we used the constraint $\vec{v} \cdot \vec{v}_1 = 0$. But, by (1),

$$(A\vec{v}) \cdot \vec{v}_1 = v \cdot (Av_1) = \lambda_1 \vec{v} \cdot \vec{v}_1 = 0,$$

so (8) implies $\mu = 0$. Plugging $\mu = 0$ into (7) yields

$$A\vec{v}_2 = \lambda \vec{v}_2,$$

and thus we have a second eigenvector, orthonormal with the first one. Repeating this process, i.e. maximizing

$$(A\vec{v}_k) \cdot \vec{v}_k,$$

subject to the constraints

$$\|\vec{v}_k\|^2 = 1, \qquad \vec{v}_k \cdot \vec{v}_1 = \cdots = v_k \cdot v_{k-1} = 0,$$

yields an orthonormal basis of eigenvectors.

(VI) For example, let

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

We observe that 0 is an eigenvalue, because the matrix has repeated rows, and the corresponding eigenspace is $\begin{bmatrix} -x_3 \\ -x_4 \\ x_3 \\ x_4 \end{bmatrix}$ which is spanned by $\begin{bmatrix} -1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$ and $\begin{bmatrix} 0 \\ -1 \\ 0 \\ 1 \end{bmatrix}$. Similarly, 2 is an eigenvalue,

[2] We could use the corresponding minimization problem to locate the smallest eigenvalue, and the other solutions to the Lagrange multiplier problem are the other eigenvalues and eigenvectors. A subtle point arises: how do we know that these maximizers and minimizers exist within the family of unit vectors? Recall that to solve some optimization problems, it is necessary to leave the family where the problem is posed in order to avoid the conclusion that there is no solution. For example, if we set out to minimize $F(x) = (x^2 - 2)^2$ over the rational numbers, to get a solution we must allow irrational numbers. If we set out to minimize the area of a rectangle subject to the constraint that the perimeter is 1, wo get a solution we must use a kind of degenerate rectangle, or line segment, with area 0. This issue does not arise for our eigenvalue problem because the real numbers are *complete* and the set of unit vectors is *compact*; see for example Section 5.1 of Shifrin's *Multivariable Mathematics* for more on this.

with eigenspace spanned by $\begin{bmatrix}1\\0\\1\\0\end{bmatrix}$ and $\begin{bmatrix}0\\1\\0\\1\end{bmatrix}$. The desired orthonormal basis is

$$\left(\frac{1}{\sqrt{2}}\begin{bmatrix}-1\\0\\1\\0\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0\\-1\\0\\1\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}1\\0\\1\\0\end{bmatrix}, \frac{1}{\sqrt{2}}\begin{bmatrix}0\\1\\0\\1\end{bmatrix}\right).$$

(VII) Another version of the Spectral Theorem says that given any $n \times n$ real symmetric matrix $A$, there exist a diagonal matrix $D$ and an invertible matrix $V$ such that $V^{-1} = V^T$ (we say that $V$ is *orthogonal*) and such that

$$A = VDV^T, \qquad D = V^T AV.$$

To derive this from the above, we let the diagonal entries of $D$ be the eigenvalues $\lambda_1, \dots \lambda_n$, and the columns of $V$ be corresponding orthonormal eigenvectors.

(VIII) In the example above, we can thus take

$$D = \begin{bmatrix}0&0&0&0\\0&0&0&0\\0&0&2&0\\0&0&0&2\end{bmatrix}, \qquad V = \frac{1}{\sqrt{2}}\begin{bmatrix}-1&0&0&1\\0&-1&1&0\\1&0&0&1\\0&1&1&0\end{bmatrix},$$

and check directly that $V^TV = VV^T = I$. Note that one of the rewards of having an orthonormal basis of eigenvectors is that computing the inverse of $V$ is almost no work.

(IX) Similarly, if $A = \begin{bmatrix}1&2\\2&4\end{bmatrix}$, since the rows are proportional we see that 0 is an eigenvalue, with corresponding unit eigenvector $\frac{1}{\sqrt{5}}\begin{bmatrix}-2\\1\end{bmatrix}$. Since the the trace of $A$ is 5, the other eigenvalue is 5, and since the matrix is symmetric the eigenvector must be perpendicular, i.e. $\frac{1}{\sqrt{5}}\begin{bmatrix}1\\2\end{bmatrix}$. Thus

$$A = VDV^T, \qquad D = V^T AV.$$

with

$$D = \begin{bmatrix}0&0\\0&5\end{bmatrix}, \qquad V = V^T = \frac{1}{\sqrt{5}}\begin{bmatrix}-2&1\\1&2\end{bmatrix}.$$

Hence

$$A^{100} = \frac{1}{5}\begin{bmatrix}-2&1\\1&2\end{bmatrix}\begin{bmatrix}0&0\\0&5^{100}\end{bmatrix}\begin{bmatrix}-2&1\\1&2\end{bmatrix} = \begin{bmatrix}-2&1\\1&2\end{bmatrix}\begin{bmatrix}0&0\\5^{99}&2\cdot5^{99}\end{bmatrix} = \begin{bmatrix}5^{99}&2\cdot5^{99}\\2\cdot5^{99}&4\cdot5^{99}\end{bmatrix}.$$