# More on the meaning of eigenvalues.

We have already seen that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are the roots of the characteristic polynomial

$$p(x) = \det(xI - M) = \prod_{i=1}^{n} (x - \lambda_i)$$

We now discuss the coefficients of $p(x)$.

Example. Suppose $M$ is a $2 \times 2$ matrix.

$$p(x) = (x - \lambda_1)(x - \lambda_2) = x^2 - (\lambda_1 + \lambda_2)x + \lambda_1\lambda_2$$

$$= x^2 - \mathrm{tr}(M)\,x + \det(M).$$

Let's generalize this!

Definition. Given a subset $S \subset \{1, \ldots, n\}$ the principal minor $M_S$ of an $n \times n$ matrix $M$ is the $|S| \times |S|$ matrix

$$M_S = \left( M_{ij} \mid i, j \in S \right)$$

Example.

$$M = \begin{pmatrix} M_{11} & \cdots & & \vdots \\ \vdots & M_{22} & & \vdots \\ \vdots & & M_{33} & \\ \cdots & \cdots & & M_{44} \end{pmatrix} \qquad M_{\{2,3\}} = \begin{pmatrix} M_{22} & M_{23} \\ M_{32} & M_{33} \end{pmatrix}$$

Theorem. The characteristic polynomial

$$p(x) = \sum_{k=0}^{n} x^{n-k} c_k (-1)^k$$

where the coefficient $c_k = \sum_{\substack{S \subset \{1, \ldots, n\} \\ |S| = k}} \det(M_S)$.

We can quickly check:

$$c_1 = \sum_{i \in \{1,\dots,n\}} \det [M_{ii}] = \operatorname{tr}(M)$$

$$c_n = \sum_{S = \{1,\dots,n\}} \det M_S = \det(M).$$

Proof. Remember that the determinant is linear in each column separately. So we can expand

$$\det(xI - M) = \sum_{\substack{S \subset \{1,\dots,n\} \\ S^c = \{1,\dots,n\} - S}} \det\left[ (xI)_{-,S^c} (-M)_{-,S} \right]$$

where the matrix in square brackets has columns with indices in $S^c$ chosen from $xI$ and columns with indices in the complementary set $S$ chosen from $-M$. Now $-M_{-,S}$ negates all $|S| = k$ of those columns, so we can write ~~this~~ ~~as~~

$$\det\left[(xI)_{-,S^c}(-M_{-,S})\right] = (-1)^{|S|}\det\left[(xI)_{-,S^c}M_{-,S}\right].$$

So now let's consider the determinant of our "hybrid" matrix.

We learned in linear algebra that one can compute determinants via "expansion by minors" where we pick ~~a~~ row or column i and write

$$\det \tilde{M} = \sum_{j=1}^{n} (-1)^{j} \det \tilde{M}(ij) \cdot \tilde{M}_{ij}$$

where $\tilde{M}(ij)$ is the matrix we get by deleting row i and column j, and $\tilde{M}_{ij}$ is the i,j th element of $\tilde{M}$.

In fact, there's a more general procedure known as "Laplace expansion".

Here we fix any $p$ rows (or columns) $\mathcal{I} = (i_1, \ldots, i_p)$ and we can write

$$\det \tilde{M} = \sum_{\substack{\mathcal{J} = \{j_1, \ldots, j_p\} \in \{1, \ldots, n\} \\ \text{(in order)}}} (-1)^{i+j} \det \tilde{M}\binom{\mathcal{I}}{\mathcal{J}} \det \tilde{M}\binom{\mathcal{I}^c}{\mathcal{J}^c}$$

where $i = i_1 + \ldots + i_p$, $j = j_1 + \ldots + j_p$. Here $\tilde{M}\binom{S_1}{S_2}$ is the submatrix of rows chosen from $S_1$ and columns chosen from $S_2$.

(Boss, right?)

Now we compute $\det\left[(xI)_{-,S^c} \; M_{-,S}\right]$ by Laplace expansion on the <u>columns</u> in $S^c$.

Since we are expanding the hybrid matrix in columns chosen from $xI$, all of the $\tilde{M}\binom{\mathcal{I}}{\mathcal{J}} = xI\binom{\mathcal{I}}{\mathcal{J}}$ have zero determinant except when $\mathcal{I} = \mathcal{J}$, in which case $\det\left(xI\binom{\mathcal{I}}{\mathcal{I}}\right) = x^{|\mathcal{I}|}$. The complementary minor $\tilde{M}\binom{\mathcal{I}^c}{\mathcal{I}^c}$ consists of ~~rows and~~ columns chosen from $M$, and is actually the principal minor $M_{\mathcal{I}^c}$.

We then have

$$\det(xI - M) = \sum_{\mathcal{I}^c \in \{1,\dots,n\}} x^{n-|\mathcal{I}_c|} (-1)^{|\mathcal{I}_c|} \det M_{\mathcal{I}_c}$$

as desired. $\square$

Since we showed in homework that the eigenvalues (and hence characteristic polynomials) of similar matrices are the same, we could change $M$ to a similar diagonal matrix $\Lambda$ and observe

Corollary. The characteristic polynomial

$$p(x) = \sum_{k=0}^{n} x^{n-k} (-1)^k c_k$$

where $c_k = \sum_{\substack{i_1 \leq \cdots \leq i_k \\ \in 1, \ldots, n}} \lambda_{i_1} \cdots \lambda_{i_k}$.

These $c_k$ (all products of $k$ eigenvalues, summed) are called elementary symmetric polynomials in the $\lambda_i$.

We have related eigenvalues to det and tr, but we still don't have a Key geometric understanding of eigenvalues.

Recall that if $A$ is a symmetric matrix, we defined $\langle \vec{x}, \vec{y} \rangle_A := \langle \vec{x}, A\vec{y} \rangle$.

Courant-Fischer Theorem.

Let ~~A~~ $M$ be a symmetric matrix with eigenvalues $\lambda_1 \geq \dots \geq \lambda_n$. Then

$$\lambda_K = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S) = K}} \ \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$$

$$= \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T) = n-K+1}} \ \max_{\substack{x \in T \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$$

where $S, T$ are subspaces of $\mathbb{R}^n$.

For instance, the largest eigenvalue

$$\lambda_1 = \max_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$$

while the smallest eigenvalue

$$\lambda_n = \min_{\substack{x \in \mathbb{R}^n \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$$

The function $\dfrac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} = \dfrac{Q_M(\vec{x})}{Q_I(\vec{x})}$ is called the Rayleigh quotient.

Proof. The proof uses some techniques we'll return to later, so it's worth doing.

Let $\vec{\psi}_1, \dots, \vec{\psi}_n$ be an orthonormal basis of eigenvectors corresponding to the eigenvalues $\lambda_1, \dots, \lambda_n$.

We first recall that any $\vec{x} \in \mathbb{R}^n$ can be written in this basis as

$$\vec{x} = \sum \langle \vec{x}, \vec{\psi}_i \rangle \vec{\psi}_i$$

Now

$$\langle \vec{x}, M\vec{x} \rangle = \sum_{i,j} \langle \langle \vec{x}, \vec{\psi}_i \rangle \vec{\psi}_i, M \langle \vec{x}, \vec{\psi}_j \rangle \vec{\psi}_j \rangle$$

$$= \sum_{i,j} \langle \vec{x}, \vec{\psi}_i \rangle \langle \vec{x}, \vec{\psi}_j \rangle \langle \vec{\psi}_i, \lambda_j \vec{\psi}_j \rangle$$

b/c $M\psi_j = \lambda_j \psi_j$.

$$= \sum_i \langle \vec{x}, \vec{\psi}_i \rangle^2 \lambda_i \quad (\text{b/c } \langle \vec{\psi}_i, \vec{\psi}_j \rangle = \delta_{ij})$$

We now show

$$\lambda_K = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim S = K}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, M\vec{x} \rangle}{\langle \vec{x}, \vec{x} \rangle}$$

First, suppose $S = \text{span}(\vec{\psi}_1, \ldots, \vec{\psi}_K)$.

Then any $\vec{x} \in S$ can be written as

$$\vec{x} = \sum_{i=1}^{K} \langle \vec{x}, \vec{\psi}_i \rangle \psi_i$$

so for these $\vec{x}$

$$\frac{\langle \vec{x}, M\vec{x} \rangle}{\langle \vec{x}, \vec{x} \rangle} = \frac{\sum_{i=1}^{K} \langle \vec{x}, \vec{\psi}_i \rangle^2 \lambda_i}{\sum_{i=1}^{K} \langle \vec{x}, \vec{\psi}_i \rangle^2}$$

but $\lambda_1 \geq \cdots \geq \lambda_K$, so this is all

$$\geq \frac{\sum_{i=1}^{K} \langle \vec{x}, \vec{\psi}_i \rangle^2 \lambda_K}{\sum_{i=1}^{K} \langle \vec{x}, \vec{\psi}_i \rangle^2} = \lambda_K.$$

Thus (for this $S$), $\displaystyle \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \geq \lambda_K$,

and so

$$\max_{\substack{S \subseteq \mathbb{R}^n \\ \dim S = K}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \geq \lambda_K$$

Now let's go the other direction. Let $T = \operatorname{span}(\vec{\Psi}_K, \vec{\Psi}_{K+1}, \ldots, \vec{\Psi}_n)$. Since $T$ has dimension $(n-K)+1$, any $S$ of $\dim(S) = K$ has an intersection with $T$ of $\dim \geq 1$.

$$\min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \leq \min_{\substack{x \in S \cap T \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \leq \max_{\substack{x \in T \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$$

But like before, any $\vec{x} \in T$ can be written $\vec{x} = \sum_{i=K}^{n} \langle \vec{x}, \vec{\Psi}_i \rangle \vec{\Psi}_i$, so

$$\frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} = \frac{\sum_{i=K}^{n} \lambda_i \langle \vec{x}, \vec{\Psi}_i \rangle^2}{\sum_{i=K}^{n} \langle \vec{x}, \vec{\Psi}_i \rangle^2} \leq \frac{\sum_{i=K}^{n} \lambda_K \langle \vec{x}, \vec{\Psi}_i \rangle^2}{\sum_{i=K}^{n} \langle \vec{x}, \vec{\Psi}_i \rangle^2} = \lambda_K.$$

Thus

$$\max_{\substack{S \subseteq \mathbb{R}^n \\ \dim S = K}} \min_{\substack{x \in S \\ x \neq 0}} \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \leq \lambda_K,$$

Completing the proof. $\square$

Now that we've introduced the Rayleigh quotient, let's see that it can be used to prove the spectral theorem. We're going to prove a weaker statement.

Theorem. Let $M$ be a symmetric matrix and $\vec{x}$ be a vector $\neq 0$ that maximizes $\frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle}$. Then $M\vec{x} = \lambda_1 \vec{x}$ where $\lambda_1$ is the largest eigenvalue of $M$.

We are going to use (multivariable) calculus! First, note that for any $k \neq 0$,

$$f(x) = \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} \text{ is equal to } f(k\vec{x}) = \frac{\langle k\vec{x}, k\vec{x} \rangle_M}{\langle k\vec{x}, k\vec{x} \rangle}.$$

So the max of $f(\vec{x})$ over $\mathbb{R}^n$ is equal to the max over $S^{n-1}$. As $S^{n-1}$ is compact and $f(\vec{x})$ is continuous, the max is reached at some $\vec{x}_0$ with $|\vec{x}_0| = 1$.

Now $f(\vec{x})$ is differentiable, so $\nabla f(\vec{x}_0) = 0$. Let's compute $\nabla f$.

$$\nabla \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} = \frac{\langle \vec{x}, \vec{x} \rangle \nabla \langle \vec{x}, \vec{x} \rangle_M - \langle \vec{x}, \vec{x} \rangle_M \nabla \langle \vec{x}, \vec{x} \rangle}{\langle \vec{x}, \vec{x} \rangle^2}$$

But

$$\nabla \langle \vec{x}, \vec{x} \rangle_M = \nabla \Big( \sum_{i,j} x_i x_j M_{ij} \Big).$$

Since $\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n}\right)$, we see

$$\nabla\Big(\sum_{i,j} x_i x_j M_{ij}\Big)_K = \sum_{i,j} \frac{\partial}{\partial x_K} x_i x_j M_{ij}$$

$$= \sum_i x_i M_{iK} + \sum_j x_j M_{Kj}$$

$$= (2M\vec{x})_K$$

NOTE! We used symmetry of $M$ here.

So we have

$$\nabla \frac{\langle \vec{x}, \vec{x} \rangle_M}{\langle \vec{x}, \vec{x} \rangle} = \frac{2\langle \vec{x}, \vec{x} \rangle M\vec{x} - 2\langle \vec{x}, M\vec{x} \rangle \vec{x}}{\langle \vec{x}, \vec{x} \rangle}$$

at $\vec{x}_0$, ~~we have~~ this gradient is zero, so

$$\langle \vec{x}_0, \vec{x}_0 \rangle M\vec{x}_0 = \cancel{2} \langle \vec{x}_0, M\vec{x}_0 \rangle \vec{x}_0$$

and

$$M\vec{x}_0 = \frac{\langle \vec{x}_0, M\vec{x}_0 \rangle}{\langle \vec{x}_0, \vec{x}_0 \rangle} \vec{x}_0 = \lambda \vec{x}_0 .$$

Now we must show that $\lambda$ is the <u>largest</u> eigenvalue. Observe that for <u>any</u> eigenvector $\vec{y}$ with eigenvalue $\mu$,

$$\frac{\langle \vec{y}, M\vec{y} \rangle}{\langle \vec{y}, \vec{y} \rangle} = \frac{\langle \vec{y}, \mu\vec{y} \rangle}{\langle \vec{y}, \vec{y} \rangle} = \mu, \text{ so } \mu = f(\vec{y}).$$

Since $x_0$ maximizes $f$ over all such $\vec{y}$, $\lambda = f(\vec{x}_0)$ must be the <u>largest</u> eigenvalue. □

We can prove the entire spectral theorem from this once we observe

Lemma. <u>If</u> $\vec{x}$ is an eigenvector of a symmetric matrix $M$ and $\langle \vec{y}, \vec{x} \rangle = 0$, then $\langle M\vec{y}, \vec{x} \rangle = 0$.

Proof. $\langle M\vec{y}, \vec{x} \rangle = \langle \vec{y}, M\vec{x} \rangle = \langle \vec{y}, \lambda\vec{x} \rangle = \lambda\langle \vec{y}, \vec{x} \rangle = 0$.

Spectral Theorem. If $M$ is an $n \times n$ real symmetric matrix, there are numbers $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n$ and orthonormal vectors $\vec{\psi}_1, \dots, \vec{\psi}_n$ so that $M\vec{\psi}_i = \lambda_i \vec{\psi}_i$.

Further,

$$\vec{\psi}_1 \in \underset{\|\vec{x}\|=1}{\arg\max} \langle \vec{x}, \vec{x} \rangle_M$$

while for $2, \dots, n$

$$\vec{\psi}_i \in \underset{\substack{|\vec{x}|=1 \\ \langle \vec{x}, \vec{\psi}_j \rangle = 0 \\ \text{for } j<i}}{\arg\max} \langle \vec{x}, \vec{x} \rangle_M$$

Recall $\arg\max f(\vec{x}) = \{ x \mid f(\vec{x}) \text{ is maximized} \}$.

A full proof is in the book, but the idea should be pretty clear: apply the previous theorem repeatedly to the smaller and smaller subspaces of $\mathbb{R}^n$ normal to the $\varphi_1, \ldots, \varphi_k$ as you build them.