Chapter 2

Eigenvalues and Optimization: The Courant-Fischer Theorem

One of the reasons that the eigenvalues of matrices have meaning is that they arise as the solution to natural optimization problems. The formal statement of this is given by the Courant-Fischer Theorem. We begin by using the Spectral Theorem to prove the Courant-Fischer Theorem. We then prove the Spectral Theorem in a form that is almost identical to Courant-Fischer.

The Rayleigh quotient of a vector $\boldsymbol{x}$ with respect to a matrix $\boldsymbol{M}$ is defined to be

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{2.1}$$

The Rayleigh quotient of an eigenvector is its eigenvalue: if $\boldsymbol{M}\boldsymbol{\psi} = \mu\boldsymbol{\psi}$, then

$$\frac{\boldsymbol{\psi}^T \boldsymbol{M} \boldsymbol{\psi}}{\boldsymbol{\psi}^T \boldsymbol{\psi}} = \frac{\boldsymbol{\psi}^T \mu \boldsymbol{\psi}}{\boldsymbol{\psi}^T \boldsymbol{\psi}} = \mu.$$

The Courant-Fischer Theorem tells us that the vectors $\boldsymbol{x}$ that maximize the Rayleigh quotient are exactly the eigenvectors of the largest eigenvalue of $\boldsymbol{M}$. In fact it supplies a similar characterization of all the eigenvalues of a symmetric matrix.

Theorem 2.0.1 (Courant-Fischer Theorem). *Let $\boldsymbol{M}$ be a symmetric matrix with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n$. Then,*

$$\mu_k = \max_{\substack{S \subseteq \mathbb{R}^n \\ \dim(S)=k}} \min_{\substack{\boldsymbol{x} \in S \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \min_{\substack{T \subseteq \mathbb{R}^n \\ \dim(T)=n-k+1}} \max_{\substack{\boldsymbol{x} \in T \\ \boldsymbol{x} \neq \boldsymbol{0}}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}},$$

where the maximization and minimization are over subspaces S and T of $\mathbb{R}^n$.

Be warned that we will often neglect to include the condition $\boldsymbol{x} \neq \boldsymbol{0}$, but we always intend it.

2.1 The First Proof

As with many proofs in Spectral Graph Theory, we begin by expanding a vector $\boldsymbol{x}$ in the basis of eigenvectors of $\boldsymbol{M}$. Let's recall how this is done.

Let $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n$ be an orthonormal basis of eigenvectors of $\boldsymbol{M}$ corresponding to $\mu_1, \dots, \mu_n$. As these are an orthonormal basis, we may write

$$\boldsymbol{x} = \sum_i c_i \boldsymbol{\psi}_i, \qquad \text{where } c_i = \boldsymbol{\psi}_i^T \boldsymbol{x}.$$

There are many ways to verify this. We let $\boldsymbol{\Psi}$ be the matrix whose columns are $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n$, and recall that the matrix $\boldsymbol{\Psi}$ is said to be *orthogonal* if its columns are orthonormal vectors. Also recall that the orthogonal matrices are exactly those matrices $\boldsymbol{\Psi}$ for which $\boldsymbol{\Psi}\boldsymbol{\Psi}^T = \boldsymbol{I}$, and that this implies that $\boldsymbol{\Psi}^T \boldsymbol{\Psi} = \boldsymbol{I}$. We now verify that

$$\sum_i c_i \boldsymbol{\psi}_i = \sum_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \boldsymbol{x} = \left(\sum_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \right) \boldsymbol{x} = \left(\boldsymbol{\Psi}\boldsymbol{\Psi}^T \right) \boldsymbol{x} = \boldsymbol{I}\boldsymbol{x} = \boldsymbol{x}.$$

When confused by orthonormal bases, just pretend that they are the basis of elementary unit vectors. For example, you know that

$$\boldsymbol{x} = \sum_i \boldsymbol{x}(i) \boldsymbol{\delta}_i, \qquad \text{and that } \boldsymbol{x}(i) = \boldsymbol{\delta}_i^T \boldsymbol{x}.$$

The first step in the proof is to express the Laplacian quadratic form of $\boldsymbol{x}$ in terms of the expansion of $\boldsymbol{x}$ in the eigenbasis.

Lemma 2.1.1. *Let $\boldsymbol{M}$ be a symmetric matrix with eigenvalues $\mu_1, \dots, \mu_n$ and a corresponding orthonormal basis of eigenvectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n$. Let $\boldsymbol{x}$ be a vector whose expansion in the eigenbasis is*

$$\boldsymbol{x} = \sum_{i=1}^n c_i \boldsymbol{\psi}_i.$$

Then,

$$\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} = \sum_{i=1}^n c_i^2 \mu_i.$$

Proof. Compute:

$$\begin{aligned}
\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} &= \left(\sum_i c_i \boldsymbol{\psi}_i\right)^T \boldsymbol{M} \left(\sum_j c_j \boldsymbol{\psi}_j\right) \\
&= \left(\sum_i c_i \boldsymbol{\psi}_i\right)^T \left(\sum_j c_j \mu_j \boldsymbol{\psi}_j\right) \\
&= \sum_{i,j} c_i c_j \mu_j \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j \\
&= \sum_i c_i^2 \mu_i,
\end{aligned}$$

as

$$\boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = \begin{cases} 0 & \text{for } i \neq j \\ 1 & \text{for } i = j. \end{cases}$$

□

Proof of 2.0.1. Let $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n$ be an orthonormal set of eigenvectors of $\boldsymbol{M}$ corresponding to $\mu_1, \dots, \mu_n$. We will just verify the first characterization of μ_k. The other is similar.

First, let's verify that μ_k is achievable. Let S be the span of $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k$. We can expand every $\boldsymbol{x} \in S$ as

$$\boldsymbol{x} = \sum_{i=1}^{k} c_i \boldsymbol{\psi}_i.$$

Applying Lemma 2.1.1 we obtain

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=1}^{k} \mu_i c_i^2}{\sum_{i=1}^{k} c_i^2} \geq \frac{\sum_{i=1}^{k} \mu_k c_i^2}{\sum_{i=1}^{k} c_i^2} = \mu_k.$$

So,

$$\min_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \geq \mu_k.$$

To show that this is in fact the maximum, we will prove that for all subspaces S of dimension k,

$$\min_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \mu_k.$$

Let T be the span of $\boldsymbol{\psi}_k, \dots, \boldsymbol{\psi}_n$. As T has dimension $n - k + 1$, every S of dimension k has an intersection with T of dimension at least 1. So,

$$\min_{\boldsymbol{x} \in S} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \min_{\boldsymbol{x} \in S \cap T} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \max_{\boldsymbol{x} \in T} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Any $\boldsymbol{x}$ in T may be expressed as

$$\boldsymbol{x} = \sum_{i=k}^{n} c_i \boldsymbol{\psi}_i,$$

and so for $\boldsymbol{x}$ in T

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{\sum_{i=k}^n \mu_i c_i^2}{\sum_{i=k}^n c_i^2} \leq \frac{\sum_{i=k}^n \mu_k c_i^2}{\sum_{i=k}^n c_i^2} = \mu_k.$$

□

2.2 Proof of the Spectral Theorem

We begin the second proof by showing that the Rayleigh quotient is maximized at an eigenvector of μ_1.

Theorem 2.2.1. *Let $\boldsymbol{M}$ be a symmetric matrix and let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient with respect to $\boldsymbol{M}$:*

$$\frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Then, $\boldsymbol{M}\boldsymbol{x} = \mu_1 \boldsymbol{x}$, where μ_1 is the largest eigenvalue of $\boldsymbol{M}$. Conversely, the minimum is achieved by eigenvectors of the smallest eigenvalue of $\boldsymbol{M}$.

Proof. We first observe that the maximum is achieved: as the Rayleigh quotient is homogeneous, it suffices to consider unit vectors $\boldsymbol{x}$. As the set of unit vectors is a closed and compact set, the maximum is achieved on this set.

Now, let $\boldsymbol{x}$ be a non-zero vector that maximizes the Rayleigh quotient. We recall that the gradient of a function at its maximum must be the zero vector. Let's compute that gradient.

We have[1]

$$\nabla \boldsymbol{x}^T \boldsymbol{x} = 2\boldsymbol{x},$$

and

$$\nabla \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} = 2\boldsymbol{M}\boldsymbol{x}.$$

So,

$$\nabla \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} = \frac{(\boldsymbol{x}^T \boldsymbol{x})(2\boldsymbol{M}\boldsymbol{x}) - (\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x})(2\boldsymbol{x})}{(\boldsymbol{x}^T \boldsymbol{x})^2}.$$

In order for this to be zero, we must have

$$(\boldsymbol{x}^T \boldsymbol{x})\boldsymbol{M}\boldsymbol{x} = (\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x})\boldsymbol{x},$$

which implies

$$\boldsymbol{M}\boldsymbol{x} = \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \boldsymbol{x}.$$

That is, if and only if $\boldsymbol{x}$ is an eigenvector of $\boldsymbol{M}$ with eigenvalue equal to its Rayleigh quotient. As $\boldsymbol{x}$ maximizes the Rayleigh quotient, this eigenvalue must be μ_1. □

[1] In case you are not used to computing gradients of functions of vectors, you can derive these directly by reasoning like

$$\frac{\partial}{\partial \boldsymbol{x}(a)} \boldsymbol{x}^T \boldsymbol{x} = \frac{\partial}{\partial \boldsymbol{x}(a)} \sum_b \boldsymbol{x}(b)^2 = 2\boldsymbol{x}(a).$$

We now prove the Spectral Theorem by generalizing this characterization to all of the eigenvalues of $\boldsymbol{M}$. The idea is to always use Theorem 2.2.1 to show that a vector is an eigenvector. To do this, we must modify the matrix for each vector.

Theorem 2.2.2. *Let $\boldsymbol{M}$ be an n-dimensional real symmetric matrix. There exist numbers $\mu_1, \dots, \mu_n$ and orthonormal vectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_n$ such that $\boldsymbol{M}\boldsymbol{\psi}_i = \mu_i \boldsymbol{\psi}_i$. Moreover,*

$$\boldsymbol{\psi}_1 \in \arg\max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x},$$

and for $2 \leq i \leq n$

$$\boldsymbol{\psi}_i \in \arg \max_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0, \text{for } j<i}} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}. \tag{2.2}$$

Similarly,

$$\boldsymbol{\psi}_i \in \arg \min_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0, \text{for } j>i}} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}.$$

Proof. We use Theorem 2.2.1 to obtain $\boldsymbol{\psi}_1$ and μ_1, and would like to proceed by induction. But first, we reduce to the case of positive definite matrices.

By Theorem 2.2.1, we also know that there is a μ_n such that

$$\mu_n = \min_{\boldsymbol{x}} \frac{\boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}^T}{\boldsymbol{x}^T \boldsymbol{x}}.$$

Now consider the matrix $\widetilde{\boldsymbol{M}} = \boldsymbol{M} + (1 - \mu_n)\boldsymbol{I}$. For all $\boldsymbol{x}$ such that $\|\boldsymbol{x}\| = 1$,

$$\boldsymbol{x}^T \widetilde{\boldsymbol{M}} \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} + 1 - \mu_n \geq 1.$$

So, $\widetilde{\boldsymbol{M}}$ is positive definite. As $\widetilde{\boldsymbol{M}}\boldsymbol{x} = \boldsymbol{M}\boldsymbol{x} + (1 - \mu_n)\boldsymbol{x}$, the eigenvectors of $\widetilde{\boldsymbol{M}}$ and $\boldsymbol{M}$ are the same. Thus it suffices to prove the theorem for positive definite matrices.

We henceforth assume without loss of generality that $\boldsymbol{M}$ is positive definite, and proceed by induction on k. Assuming that we have eigenvectors $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k$ satisfying (2.2), we construct $\boldsymbol{\psi}_{k+1}$. Define

$$\boldsymbol{M}_k = \boldsymbol{M} - \sum_{i=1}^{k} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T.$$

For $j \leq k$ we have

$$\boldsymbol{M}_k \boldsymbol{\psi}_j = \boldsymbol{M} \boldsymbol{\psi}_j - \sum_{i=1}^{k} \mu_i \boldsymbol{\psi}_i \boldsymbol{\psi}_i^T \boldsymbol{\psi}_j = \mu_j \boldsymbol{\psi}_j - \mu_j \boldsymbol{\psi}_j = \boldsymbol{0}.$$

So, for vectors $\boldsymbol{x}$ that are orthogonal to $\boldsymbol{\psi}_1, \dots, \boldsymbol{\psi}_k$,

$$\boldsymbol{M}_k \boldsymbol{x} = \boldsymbol{M} \boldsymbol{x}, \quad \boldsymbol{x}^T \boldsymbol{M}_k \boldsymbol{x} = \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}, \quad \text{and}$$

$$\arg \max_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0, \text{for } j \leq k}} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x} \leq \arg \max_{\|\boldsymbol{x}\|=1} \boldsymbol{x}^T \boldsymbol{M}_k \boldsymbol{x}. \tag{2.3}$$

Now, let $\boldsymbol{y}$ be a unit vector that maximizes $\boldsymbol{y}^T \boldsymbol{M}_k \boldsymbol{y}$. We know from Theorem 2.2.1 that $\boldsymbol{y}$ is an eigenvector of $\boldsymbol{M}_k$. Call its eigenvalue μ. We now show that $\boldsymbol{y}$ must be orthogonal to each of $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$. Let

$$\widetilde{\boldsymbol{y}} = \boldsymbol{y} - \sum_{i=1}^{k} \boldsymbol{\psi}_i (\boldsymbol{\psi}_i^T \boldsymbol{y})$$

be the projection of $\boldsymbol{y}$ orthogonal to $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$, and let $\widehat{\boldsymbol{y}} = \widetilde{\boldsymbol{y}} / \|\widetilde{\boldsymbol{y}}\|$. As $\boldsymbol{M}_k \boldsymbol{\psi}_i = \boldsymbol{0}$ for $i \leq k$, we know that $\widetilde{\boldsymbol{y}}^T \boldsymbol{M}_k \widetilde{\boldsymbol{y}} = \boldsymbol{y}^T \boldsymbol{M} \boldsymbol{y}$. If $\boldsymbol{y}$ is not orthogonal to these vectors, that is if some $\boldsymbol{\psi}_i^T \boldsymbol{x}$ is nonzero, then $\|\widetilde{\boldsymbol{y}}\| < \|\boldsymbol{y}\|$. As $\widetilde{\boldsymbol{y}}^T \boldsymbol{M} \widetilde{\boldsymbol{y}} > 0$, this would imply that for the unit vector $\widehat{\boldsymbol{y}}$, $\widehat{\boldsymbol{y}}^T \boldsymbol{M}_k \widehat{\boldsymbol{y}} > \boldsymbol{y}^T \boldsymbol{M}_k \boldsymbol{y}$, a contradiction. As $\boldsymbol{y}$ is orthogonal to $\boldsymbol{\psi}_1, \ldots, \boldsymbol{\psi}_k$ and it is an eigenvector of $\boldsymbol{M}_k$, it is also an eigenvector of $\boldsymbol{M}$:

$$\boldsymbol{M} \boldsymbol{y} = \boldsymbol{M}_k \boldsymbol{y} = \mu \boldsymbol{y},$$

and by (2.3)

$$\boldsymbol{y} \in \arg \max_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{\psi}_j = 0, \text{for } j \leq k}} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x}.$$

We now set $\boldsymbol{\psi}_{k+1} = \boldsymbol{y}$ and $\mu_{k+1} = \mu$. □

2.3 Notes

The characterization of eigenvalues by maximizing or minimizing the Rayleigh quotient only works for symmetric matrices. The analogous quantities for non-symmetric matrices $\boldsymbol{A}$ are the *singular vectors* and *singular values* of $\boldsymbol{A}$, which are the eigenvectors of $\boldsymbol{A}\boldsymbol{A}^T$ and $\boldsymbol{A}^T \boldsymbol{A}$, and the square roots of the eigenvalues of those matrices.

2.4 Exercise

1. A tighter characterization.

Tighten Theorem 2.2.2 by proving that for every sequence of vectors $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ such that

$$\boldsymbol{x}_i \in \arg \max_{\substack{\|\boldsymbol{x}\|=1 \\ \boldsymbol{x}^T \boldsymbol{x}_j = 0, \text{for } j < i}} \boldsymbol{x}^T \boldsymbol{M} \boldsymbol{x},$$

each $\boldsymbol{x}_i$ is an eigenvector of $\boldsymbol{M}$.