Appendix A

Rayleigh Ratios and the Courant-Fischer Theorem

The most important property of symmetric matrices is that they have real eigenvalues and that they can be diagonalized with respect to an orthogonal matrix.

Thus, if $A$ is an $n \times n$ symmetric matrix, then it has $n$ real eigenvalues $\lambda_1, \ldots, \lambda_n$ (not necessarily distinct), and there is an orthonormal basis of eigenvectors $(u_1, \ldots, u_n)$ (for a proof, see Gallier [6]).
Another fact that is used frequently in optimization problem is that the eigenvalues of a symmetric matrix are characterized in terms of what is known as the Rayleigh ratio, defined by

$$R(A)(x) = \frac{x^\top Ax}{x^\top x}, \quad x \in \mathbb{R}^n, x \neq 0.$$

The following proposition is often used to prove the correctness of various optimization or approximation problems (for example PCA).
Proposition A.1. (Rayleigh–Ritz) If $A$ is a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and if $(u_1, \ldots, u_n)$ is any orthonormal basis of eigenvectors of $A$, where $u_i$ is a unit eigenvector associated with $\lambda_i$, then

$$\max_{x \neq 0} \frac{x^\top Ax}{x^\top x} = \lambda_n$$

(with the maximum attained for $x = u_n$), and

$$\max_{x \neq 0, x \in \{u_{n-k+1}, \ldots, u_n\} \perp} \frac{x^\top Ax}{x^\top x} = \lambda_{n-k}$$

(with the maximum attained for $x = u_{n-k}$), where $1 \leq k \leq n - 1$.

Equivalently, if $V_k$ is the subspace spanned by $(u_1, \ldots, u_k)$, then

$$\lambda_k = \max_{x \neq 0, x \in V_k} \frac{x^\top Ax}{x^\top x}, \quad k = 1, \ldots, n.$$
For our purposes, we also need the version of Proposition A.1 applying to \( \min \) instead of \( \max \).

**Proposition A.2.** (Rayleigh–Ritz) If \( A \) is a symmetric \( n \times n \) matrix with eigenvalues \( \lambda_1 \leq \lambda_2 \leq \ldots \leq \lambda_n \) and if \( (u_1, \ldots, u_n) \) is any orthonormal basis of eigenvectors of \( A \), where \( u_i \) is a unit eigenvector associated with \( \lambda_i \), then

\[
\min_{x \neq 0} \frac{x^\top A x}{x^\top x} = \lambda_1
\]

(with the minimum attained for \( x = u_1 \)), and

\[
\min_{x \neq 0, x \in \{u_1, \ldots, u_{i-1}\}^\perp} \frac{x^\top A x}{x^\top x} = \lambda_i
\]

(with the minimum attained for \( x = u_i \)), where \( 2 \leq i \leq n \).

Equivalently, if \( W_k = V_{k-1}^\perp \) denotes the subspace spanned by \( (u_k, \ldots, u_n) \) (with \( V_0 = (0) \)), then

\[
\lambda_k = \min_{x \neq 0, x \in W_k} \frac{x^\top A x}{x^\top x} = \min_{x \neq 0, x \in V_{k-1}^\perp} \frac{x^\top A x}{x^\top x}, \quad k = 1, \ldots, n.
\]
Propositions A.1 and A.2 together are known as the 
*Rayleigh–Ritz theorem*.

As an application of Propositions A.1 and A.2, we give a proof of a proposition which is the key to the proof of Theorem 2.2.

Given an $n \times n$ symmetric matrix $A$ and an $m \times m$ symmetric $B$, with $m \leq n$, if $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of $A$ and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m$ are the eigenvalues of $B$, then we say that the *eigenvalues of $B$ interlace the eigenvalues of $A$* if

$$\lambda_i \leq \mu_i \leq \lambda_{n-m+i}, \quad i = 1, \ldots, m.$$ 

The following proposition is known as the *Poincaré separation theorem*. 
**Proposition A.3.** Let $A$ be an $n \times n$ symmetric matrix, $R$ be an $n \times m$ matrix such that $R^\top R = I$ (with $m \leq n$), and let $B = R^\top AR$ (an $m \times m$ matrix). The following properties hold:

(a) The eigenvalues of $B$ interlace the eigenvalues of $A$.

(b) If $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ are the eigenvalues of $A$ and $\mu_1 \leq \mu_2 \leq \cdots \leq \mu_m$ are the eigenvalues of $B$, and if $\lambda_i = \mu_i$, then there is an eigenvector $v$ of $B$ with eigenvalue $\mu_i$ such that $Rv$ is an eigenvector of $A$ with eigenvalue $\lambda_i$.

Observe that Proposition A.3 implies that

$$\lambda_1 + \cdots + \lambda_m \leq \text{tr}(R^\top AR) \leq \lambda_{n-m+1} + \cdots + \lambda_n.$$

The left inequality is used to prove Theorem 2.2.
For the sake of completeness, we also prove the Courant–Fischer characterization of the eigenvalues of a symmetric matrix.

Theorem A.4. (Courant–Fischer) Let $A$ be a symmetric $n \times n$ matrix with eigenvalues $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and let $(u_1, \ldots, u_n)$ be any orthonormal basis of eigenvectors of $A$, where $u_i$ is a unit eigenvector associated with $\lambda_i$. If $\mathcal{V}_k$ denotes the set of subspaces of $\mathbb{R}^n$ of dimension $k$, then

$$
\lambda_k = \max_{W \in \mathcal{V}_{n-k+1}} \min_{x \in W, x \neq 0} \frac{x^\top Ax}{x^\top x},
$$

$$
\lambda_k = \min_{W \in \mathcal{V}_k} \max_{x \in W, x \neq 0} \frac{x^\top Ax}{x^\top x}.
$$
Bibliography


