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

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

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 λ_i, then*

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

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

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

(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

$$\lambda_k = \max_{x \neq 0, x \in V_k} \frac{x^\top A x}{x^\top x}, \quad k = 1, \ldots, n.$$

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 $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_n$ and if $(u_1, \ldots, u_n)$ is any orthonormal basis of eigenvectors of A, where u_i is a unit eigenvector associated with λ_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 μ_i such that Rv is an eigenvector of A with eigenvalue λ_i.

Observe that Proposition A.3 implies that

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

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 λ_i. If $\mathcal{V}_k$ denotes the set of subspaces of $\mathbb{R}^n$ of dimension k, then*

$$\begin{aligned}\lambda_k &= \max_{W \in \mathcal{V}_{n-k+1}} \min_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}\\ \lambda_k &= \min_{W \in \mathcal{V}_k} \max_{x \in W, x \neq 0} \frac{x^\top A x}{x^\top x}.\end{aligned}$$

Bibliography

[1] Nikhil Bansal, Avrim Blum, and Shuchi Chawla. Correlation clustering. *Machine Learning*, 56:89–113, 2004.

[2] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural Computation*, 15:1373–1396, 2003.

[3] Fan R. K. Chung. *Spectral Graph Theory*, volume 92 of *Regional Conference Series in Mathematics*. AMS, first edition, 1997.

[4] Eric D. Demaine and Nicole Immorlica. Correlation clustering with partial information. In S. Arora et al., editor, *Working Notes of the 6th International Workshop on Approximation Algorithms for Combinatorial Problems*, LNCS Vol. 2764, pages 1–13. Springer, 2003.

[5] Jean H. Gallier. *Discrete Mathematics*. Universitext. Springer Verlag, first edition, 2011.

[6] Jean H. Gallier. *Geometric Methods and Applications, For Computer Science and Engineering.* TAM, Vol. 38. Springer, second edition, 2011.

[7] Chris Godsil and Gordon Royle. *Algebraic Graph Theory*. GTM No. 207. Springer Verlag, first edition, 2001.

[8] H. Golub, Gene and F. Van Loan, Charles. *Matrix Computations*. The Johns Hopkins University Press, third edition, 1996.

[9] Frank Harary. On the notion of balance of a signed graph. *Michigan Math. J.*, 2(2):143–146, 1953.

[10] Jao Ping Hou. Bounds for the least laplacian eigenvalue of a signed graph. *Acta Mathematica Sinica*, 21(4):955–960, 2005.

[11] Ravikrishna Kolluri, Jonathan R. Shewchuk, and James F. O'Brien. Spectral surface reconstruction from noisy point clouds. In *Symposium on Geometry Processing*, pages 11–21. ACM Press, July 2004.

[12] Jérôme Kunegis, Stephan Schmidt, Andreas Lommatzsch, Jürgen Lerner, Ernesto William De Luca, and Sahin Albayrak. Spectral analysis of signed graphs for clustering, prediction and visualization. In *SDM'10*, pages 559–559, 2010.

[13] Jianbo Shi and Jitendra Malik. Normalized cuts and image segmentation. *Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, 2000.

[14] Daniel Spielman. Spectral graph theory. In Uwe Naumannn and Olaf Schenk, editors, *Combinatorial Scientific Computing*. CRC Press, 2012.

[15] von Luxburg Ulrike. A tutorial on spectral clustering. *Statistics and Computing*, 17(4):395–416, 2007.

[16] Stella X. Yu. *Computational Models of Perceptual Organization*. PhD thesis, Carnegie Mellon University, Pittsburgh, PA 15213, USA, 2003. Dissertation.

[17] Stella X. Yu and Jianbo Shi. Multiclass spectral clustering. In *9th International Conference on Computer Vision, Nice, France, October 13-16*. IEEE, 2003.