Extremal Properties of Eigenvalues

Seungjin Choi
Department of Computer Science
POSTECH, Korea
seungjin@postech.ac.kr

Definition 1 *Let $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ be symmetric. The Rayleigh quotient $R(\boldsymbol{x}, \boldsymbol{A})$ is defined by*

$$R(\boldsymbol{x}, \boldsymbol{A}) = \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{1}$$

Theorem 1 *Let $\boldsymbol{A} \in \mathbb{R}^{m \times m}$ be symmetric with its eigenvalues being $\{\lambda_1 \geq \cdots \geq \lambda_m\}$. For $\forall \boldsymbol{x} \neq \boldsymbol{0} \in \mathbb{R}^m$, we have*

$$\lambda_m \leq \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \leq \lambda_1, \tag{2}$$

and in particular,

$$\lambda_m = \min_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}, \tag{3}$$

$$\lambda_1 = \max_{\boldsymbol{x} \neq 0} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}}. \tag{4}$$

Proof: The eigen-decomposition of the matrix $\boldsymbol{A}$ is given by

$$\boldsymbol{A} = \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{U}^T, \tag{5}$$

where

$$\begin{aligned} \boldsymbol{U} &= [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m] \\ \boldsymbol{\Sigma} &= \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix}. \end{aligned}$$

If $\boldsymbol{y} = \boldsymbol{U}^T \boldsymbol{x}$, then we have

$$\begin{aligned} \frac{\boldsymbol{x}^T \boldsymbol{A} \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} &= \frac{x^T \boldsymbol{U} \boldsymbol{\Sigma} \boldsymbol{U}^T \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{U} \boldsymbol{U}^T \boldsymbol{x}} \\ &= \frac{\boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y}}{\boldsymbol{y}^T \boldsymbol{y}} \\ &= \frac{\sum_{i=1}^m \lambda_i y_i^2}{\sum_{i=1}^m y_i^2}. \end{aligned} \tag{6}$$

It follows from (6) that

$$\lambda_m \sum_{i=1}^{m} y_i^2 \leq \sum_{i=1}^{m} \lambda_i y_i^2 \leq \lambda_1 \sum_{i=1}^{m} y_i^2.$$

Eqs. (3) and (4) are verified by choices of $\boldsymbol{x}$ for which the bounds in (2) are attained. For instance, the lower bound is attained with $\boldsymbol{x} = \boldsymbol{x}_m$, while the upper bound holds with $\boldsymbol{x} = \boldsymbol{x}_1$. QED.

Note: For $\forall \boldsymbol{x} \neq 0$, $\boldsymbol{z} = (\boldsymbol{x}^T\boldsymbol{x})^{-\frac{1}{2}}\boldsymbol{x}$ is a unit vector. Then we have

$$\begin{aligned} \lambda_m &= \min_{\boldsymbol{z}^T\boldsymbol{z}=1} \boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z}, \\ \lambda_1 &= \max_{\boldsymbol{z}^T\boldsymbol{z}=1} \boldsymbol{z}^T\boldsymbol{A}\boldsymbol{z}. \end{aligned} \tag{7}$$

Theorem 2 *Let $\boldsymbol{A} \in \mathbb{R}^{m\times m}$ be a symmetric matrix having eigenvalues $\lambda_1 \geq \cdots \geq \lambda_m$, with $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ being a corresponding set of orthonormal eigenvectors. For $h = 1, \ldots, m$, define $\mathcal{S}_h$ and $\mathcal{T}_h$ to the vector spaces spanned by the columns of $\boldsymbol{X}_h = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_h]$ and $\boldsymbol{Y}_h = [\boldsymbol{x}_h, \ldots, \boldsymbol{x}_m]$, respectively. Then,*

$$\lambda_h = \min_{\boldsymbol{x}\in\mathcal{S}_h} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \min_{\boldsymbol{Y}_{h+1}^T\boldsymbol{x}=\boldsymbol{0}} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}, \tag{8}$$

and

$$\lambda_h = \max_{\boldsymbol{x}\in\mathcal{T}_h} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \max_{\boldsymbol{X}_{h-1}^T\boldsymbol{x}=\boldsymbol{0}} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}}, \tag{9}$$

where the vector $\boldsymbol{x} = \boldsymbol{0}$ has been excluded from the maximization and minimization processes.

Proof: We will prove the result concerning the minimum; the proof for the maximum is similar. Let $\boldsymbol{X} = [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m]$ and $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1, \ldots, \lambda_m\}$. Since $\boldsymbol{X}^T\boldsymbol{A}\boldsymbol{X} = \boldsymbol{\Lambda}$ and $\boldsymbol{X}^T\boldsymbol{X} = \boldsymbol{I}_m$, it follows that $\boldsymbol{X}_h^T\boldsymbol{X}_h = \boldsymbol{I}_h$ and $\boldsymbol{X}_h^T\boldsymbol{A}\boldsymbol{X}_h = \boldsymbol{\Lambda}_h$, where $\boldsymbol{\Lambda}_h = \mathrm{diag}\{\lambda_1, \ldots, \lambda_h\}$. Note that $\boldsymbol{x} \in \mathcal{S}_h$ if and only if there exists an $\boldsymbol{y} \in \mathbb{R}^h$ such that $\boldsymbol{x} = \boldsymbol{X}_h\boldsymbol{y}$. Then

$$\min_{\boldsymbol{x}\in\mathcal{S}_h} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{x}} = \min_{\boldsymbol{y}\neq\boldsymbol{0}} \frac{\boldsymbol{y}^T\boldsymbol{X}_h^T\boldsymbol{A}\boldsymbol{X}_h\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{X}_h^T\boldsymbol{X}_h\boldsymbol{y}} = \min_{\boldsymbol{y}\neq\boldsymbol{0}} \frac{\boldsymbol{y}^T\boldsymbol{\Lambda}_h\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}} = \lambda_h, \tag{10}$$

where the last equality follows from Theorem 1. The second version of the minimization follows immediately from the first and the fact that the null space of $\boldsymbol{Y}_{h+1}^T$ is $\mathcal{S}_h$. QED.

Remarks

- $\boldsymbol{x} \in \mathcal{S}_h$ iff $\boldsymbol{Y}_{h+1}^T\boldsymbol{x} = 0$ ($\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Y}_{h+1}^T)$).
- Since $\mathcal{S}_h \perp \mathcal{T}_h$, $\boldsymbol{x} \in \mathcal{S}_h$ implies $\boldsymbol{x} \in \mathcal{N}(\boldsymbol{Y}_{h+1}^T)$.

Theorem 3 *Let $\boldsymbol{A}$ and $\boldsymbol{B}$ be $m \times m$ matrices, with $\boldsymbol{A}$ being nonnegative definite and $\boldsymbol{B}$ positive definite. For $h = 1, \ldots, m$, define*

$$\begin{aligned} \boldsymbol{X}_h &= [\boldsymbol{x}_1, \ldots, \boldsymbol{x}_h], \\ \boldsymbol{Y}_h &= [\boldsymbol{x}_h, \ldots, \boldsymbol{x}_m], \end{aligned}$$

where $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_m$ *are linear independent eigenvectors of* $\boldsymbol{B}^{-1}\boldsymbol{A}$ *corresponding to the eigenvalues*

$$\lambda_1\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) \geq \cdots \geq \lambda_m\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right). \tag{11}$$

Then

$$\lambda_h\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) = \min_{\boldsymbol{Y}_{h+1}^T \boldsymbol{B}\boldsymbol{x}=0} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}}, \tag{12}$$

and

$$\lambda_h\left(\boldsymbol{B}^{-1}\boldsymbol{A}\right) = \max_{\boldsymbol{X}_{h-1}^T \boldsymbol{B}\boldsymbol{x}=0} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}}, \tag{13}$$

where $\boldsymbol{x} = 0$ *is excluded, and the min and max are over all* $\boldsymbol{x} \neq 0$ *when* $h = m$ *and* $h = 1$*, respectively.*

Proof: The spectral decomposition of $\boldsymbol{B}$ is $\boldsymbol{B} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{U}^T$. If we let $\boldsymbol{T} = \boldsymbol{U}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{U}^T$, then $\boldsymbol{B} = \boldsymbol{T}\boldsymbol{T} = \boldsymbol{T}^2$. Note that $\boldsymbol{T}$, like $\boldsymbol{B}$, is symmetric and nonsingular. Putting $\boldsymbol{y} = \boldsymbol{T}\boldsymbol{x}$, we find that

$$\begin{aligned}
\min_{\boldsymbol{Y}_{h+1}^T \boldsymbol{B}\boldsymbol{x}=0} \frac{\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{B}\boldsymbol{x}} &= \min_{\boldsymbol{Y}_{h+1}^T \boldsymbol{T}\boldsymbol{T}\boldsymbol{x}=0} \frac{\boldsymbol{x}^T\boldsymbol{T}\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{x}}{\boldsymbol{x}^T\boldsymbol{T}\boldsymbol{T}\boldsymbol{x}} \\
&= \min_{\boldsymbol{Y}_{h+1}^T \boldsymbol{T}\boldsymbol{y}=0} \frac{\boldsymbol{y}^T\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1}\boldsymbol{y}}{\boldsymbol{y}^T\boldsymbol{y}}.
\end{aligned} \tag{14}$$

Note that if we write $\lambda_i = \lambda_i(\boldsymbol{B}^{-1}\boldsymbol{A})$, then $\boldsymbol{B}^{-1}\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i$, so that

$$\boldsymbol{T}^{-1}\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i, \tag{15}$$

which implies

$$\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1}\boldsymbol{T}\boldsymbol{x}_i = \lambda_i\boldsymbol{T}\boldsymbol{x}_i. \tag{16}$$

Thus, $\boldsymbol{T}\boldsymbol{x}_i$ is an eigenvector of $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1}$ corresponding to the eigenvalue $\lambda_i = \lambda_i(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1})$. That is, the eigenvalues of $\boldsymbol{B}^{-1}\boldsymbol{A}$ are the same as those of $\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1}$. Since the rows of $\boldsymbol{Y}_{h+1}^T\boldsymbol{T}$ are the transpose of the eigenvectors $\boldsymbol{T}\boldsymbol{x}_{h+1}, \ldots, \boldsymbol{T}\boldsymbol{x}_m$, it follows from Theorem 2 that (14) equals $\lambda_h(\boldsymbol{T}^{-1}\boldsymbol{A}\boldsymbol{T}^{-1})$, which we have already established as being the same as $\lambda_h(\boldsymbol{B}^{-1}\boldsymbol{A})$.