San José State University

Math 250: Mathematical Data Visualization

Rayleigh Quotients

Dr. Guangliang Chen

Outline

- Ordinary Rayleigh quotients
- Generalized Rayleigh quotients
- Applications

The ordinary Rayleigh quotients

Rayleigh quotients are encountered in many statistical and machine learning problems. It is thus necessary to study it systematically.

Def 0.1. The Rayleigh quotient for a given symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ is a multivariate function $f : \mathbb{R}^n - \{\mathbf{0}\} \longmapsto \mathbb{R}$ defined by

$$f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{x}}, \quad \mathbf{x} \neq \mathbf{0}.$$

Remark. A Rayleigh quotient is always scaling invariant, that is, for any nonzero vector $\mathbf{x} \in \mathbb{R}^n$,

$$f(k\mathbf{x}) = \frac{(k\mathbf{x})^T\mathbf{A}(k\mathbf{x})}{(k\mathbf{x})^T(k\mathbf{x})} = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = f(\mathbf{x}), \qquad \text{for all } k \neq 0$$

Another way to see it is to rewrite the Rayleigh quotient as follows:

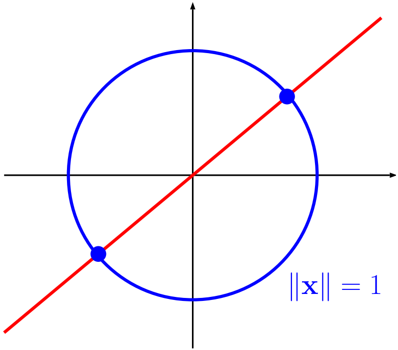
$$f(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\|\mathbf{x}\|^2} = \left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right)^T \mathbf{A} \left(\frac{\mathbf{x}}{\|\mathbf{x}\|}\right), \quad \mathbf{x} \neq \mathbf{0}.$$

It is thus enough to focus on the unit sphere in $\mathbb{R}^n$

$$S_n = \{\mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| = 1\},$$

on which the Rayleigh quotient reduces to

$$f\,|_{S_n}(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x},\ \ \mathbf{x} \in S_n.$$



Interpretation:

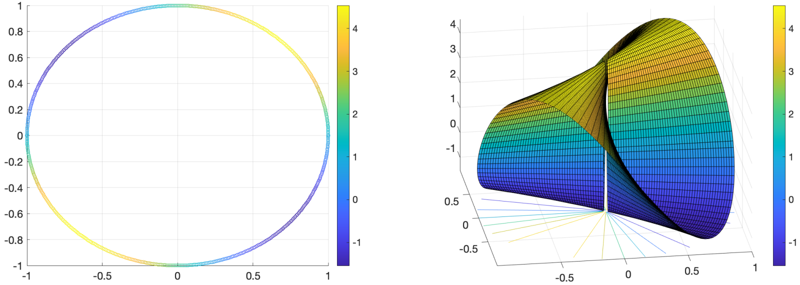
The Rayleigh quotient is essentially a quadratic form over unit sphere.

Example 0.1. The Rayleigh quotient for $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix} \in S^2(\mathbb{R})$ is

$$f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \frac{x_1^2 + 2x_2^2 + 6x_1 x_2}{x_1^2 + x_2^2}, \quad \mathbf{x} \neq \mathbf{0}.$$

It is a function defined over $\mathbb{R}^2$ with the origin excluded.

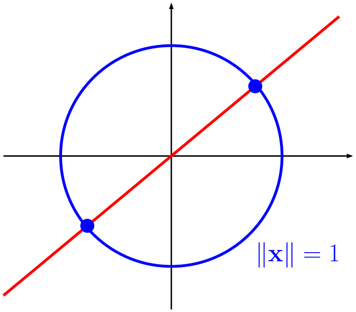
We plot below the values of f along the circle $\mathbf{x}^T\mathbf{x} = x_1^2 + x_2^2 = 1$ (left) and also the full graph in 3 dimensions (right).



Optimization of Rayleigh quotients

Problem. Given $\mathbf{A} \in S^n(\mathbb{R})$, find the maximum (or minimum) of the associated Rayleigh quotient

$$\max_{\mathbf{x} \neq \mathbf{0} \in \mathbb{R}^n} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} \quad \leftarrow \text{scaling invariant}$$



Equivalent formulations:

$$\max_{\mathbf{x} \in \mathbb{R}^n :\, \|\mathbf{x}\|=1} \mathbf{x}^T \mathbf{A} \mathbf{x}$$

$$\max_{\mathbf{x} \in \mathbb{R}^n} \mathbf{x}^T \mathbf{A} \mathbf{x} \quad \text{subject to } \|\mathbf{x}\|^2 = 1 \quad \longleftarrow \text{Constrained optimization}$$

Theorem 0.1. For any given symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$, let its largest and smallest eigenvalues be λ_1 and λ_n, with associated eigenvectors $\mathbf{v}_1, \mathbf{v}_n \in \mathbb{R}^n$, respectively. Then the maximum (or minimum) value of the associated Rayleigh quotient $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is equal to the largest (or smallest) eigenvalue of $\mathbf{A}$, achieved by the corresponding eigenvectors:

$$\max_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_1, \quad @\ \mathbf{x} = \pm\mathbf{v}_1$$

$$\min_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_n, \quad @\ \mathbf{x} = \pm\mathbf{v}_n$$

Remark. Any nonzero scalar multiple of the top (bottom) eigenvector is also a maximizer (minimizer). For simplicity, we focus on the unit-norm eigenvectors as maximizer and minimizers.

Example 0.2. For the PSD matrix $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, we have previously obtained its eigenvalues and eigenvectors

$$\lambda_1 = 5,\ \lambda_2 = 0; \quad \mathbf{v}_1 = \frac{1}{\sqrt{5}}(1,2)^T,\ \mathbf{v}_2 = \frac{1}{\sqrt{5}}(-2,1)^T$$

The associated Rayleigh quotient $Q(\mathbf{x}) = \frac{x_1^2+4x_2^2+4x_1x_2}{x_1^2+x_2^2}$ has the following extreme values:

- The maximum value of $Q(\mathbf{x})$ is $\lambda_1 = 5$, achieved at $\mathbf{x} = \pm\mathbf{v}_1$;
- The minimum is $\lambda_1 = 0$, achieved at $\mathbf{x} = \pm\mathbf{v}_2$.

The overall range of the Rayleigh quotient is thus $[0, 5]$.

Linear algebra approach

Proof. Let $\mathbf{A} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^T$ be the spectral decomposition, where $\mathbf{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ is orthogonal and $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is diagonal with sorted diagonals from large to small. Then for any unit vector $\mathbf{x}$,

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \mathbf{x}^T(\mathbf{V}\mathbf{\Lambda}\mathbf{V}^T)\mathbf{x} = (\mathbf{x}^T\mathbf{V})\mathbf{\Lambda}(\mathbf{V}^T\mathbf{x}) = \mathbf{y}^T\mathbf{\Lambda}\mathbf{y}$$

where $\mathbf{y} = \mathbf{V}^T\mathbf{x}$ is also a unit vector:

$$\|\mathbf{y}\|^2 = \mathbf{y}^T\mathbf{y} = (\mathbf{V}^T\mathbf{x})^T(\mathbf{V}^T\mathbf{x}) = \mathbf{x}^T\mathbf{V}\mathbf{V}^T\mathbf{x} = \mathbf{x}^T\mathbf{x} = 1.$$

So the original optimization problem becomes the following one:

$$\max_{\mathbf{y}\in\mathbb{R}^n:\, \|\mathbf{y}\|=1} \mathbf{y}^T \underbrace{\mathbf{\Lambda}}_{\text{diagonal}} \mathbf{y}$$

To solve this new problem, write $\mathbf{y} = (y_1, \ldots, y_n)^T$. It follows that

$$\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \sum_{i=1}^{n} \underbrace{\lambda_i}_{\text{fixed}} y_i^2 \qquad (\text{subject to } y_1^2 + y_2^2 + \cdots + y_n^2 = 1)$$

Because $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$, when $y_1^2 = 1, y_2^2 = \cdots = y_n^2 = 0$ (i.e., $\mathbf{y} = \pm\mathbf{e}_1$), the quadratic form attains its maximum value $\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \lambda_1$.

In terms of the original variable $\mathbf{x}$, the maximizer is

$$\mathbf{x} = \mathbf{V}\mathbf{y} = \mathbf{V}(\pm\mathbf{e}_1) = \pm\mathbf{v}_1.$$

Multivariable calculus approach

Proof. Alternatively, we can use the method of Lagrange multipliers to prove the theorem.

First, we form the Lagrangian function

$$L(\mathbf{x}, \lambda) = \mathbf{x}^T \mathbf{A}\mathbf{x} - \lambda(\|\mathbf{x}\|^2 - 1).$$

Next, we need to compute the partial derivatives, $\frac{\partial L}{\partial \mathbf{x}} = (\frac{\partial L}{\partial x_1}, \dots, \frac{\partial L}{\partial x_n})^T$ and $\frac{\partial L}{\partial \lambda}$, and set them equal to zero (in order to find its critical points).

For this goal, we need to know how to differentiate functions like $\mathbf{x}^T\mathbf{A}\mathbf{x}$, $\|\mathbf{x}\|^2$ with respect to the vector-valued variable $\mathbf{x}$.

We present a few formulas of such kind below (the proofs can be found in the notes).

Proposition 0.2. For any fixed symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$, matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$ and vector $\mathbf{a} \in \mathbb{R}^n$, we have

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{a}^T\mathbf{x}) = \mathbf{a}, \qquad \frac{\partial}{\partial \mathbf{x}}(\|\mathbf{x}\|^2) = 2\mathbf{x}$$

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = 2\mathbf{A}\mathbf{x}, \qquad \frac{\partial}{\partial \mathbf{x}}(\|\mathbf{B}\mathbf{x}\|^2) = 2\mathbf{B}^T\mathbf{B}\mathbf{x}$$

Now, applying the formulas obtained previously, we have

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - \lambda(2\mathbf{x}) = 0 \qquad \longrightarrow \qquad \mathbf{A}\mathbf{x} = \lambda\mathbf{x}$$
$$\frac{\partial L}{\partial \lambda} = -(\|\mathbf{x}\|^2 - 1) = 0 \qquad \longrightarrow \qquad \|\mathbf{x}\|^2 = 1$$

This implies that $\mathbf{x}, \lambda$ must be a (normalized) eigenpair of $\mathbf{A}$. For any solution $\lambda = \lambda_i, \mathbf{x} = \mathbf{v}_i$, the objective function $\mathbf{x}^T\mathbf{A}\mathbf{x}$ takes the value

$$\mathbf{v}_i^T\mathbf{A}\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i\mathbf{v}_i) = \lambda_i\|\mathbf{v}_i\|^2 = \lambda_i.$$

Therefore, the eigenvector $\mathbf{v}_1$ (corresponding to largest eigenvalue λ_1 of $\mathbf{A}$) is a global maximizer, and it yields the absolute maximum value λ_1. Similarly, the eigenvector $\mathbf{v}_n$ corresponding to the smallest eigenvalue λ_n is a global minimizer with absolute minimum λ_n.

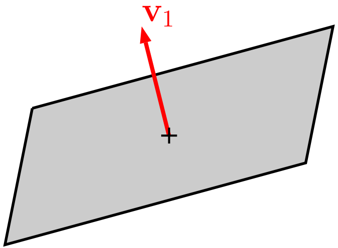
Restricted Rayleigh quotients

Sometimes, we may choose to "exclude" the top (bottom) few eigenvectors from the optimization domain when maximizing (minimizing) a Rayleigh quotient:

$$\max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{v}_1^T\mathbf{x}=0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

$$\max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{v}_1^T\mathbf{x}=\mathbf{v}_2^T\mathbf{x}=0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

In such cases, the effective domain is the orthogonal complement of the excluded eigenvector(s).



It turns out that the next eigenvector will be optimal.

Theorem 0.3 (Rayleigh-Ritz). Given a symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$, let $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ be its eigenvalues and $\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_n \in \mathbb{R}^n$ a collection of corresponding eigenvectors (in unit norm). We have

$$\max_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{v}_1^T \mathbf{x} = 0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_2 \quad (\text{when } \mathbf{x} = \pm \mathbf{v}_2)$$

$$\max_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{v}_1^T \mathbf{x} = \mathbf{v}_2^T \mathbf{x} = 0}} \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}} = \lambda_3 \quad (\text{when } \mathbf{x} = \pm \mathbf{v}_3)$$

and so on.

Example 0.3. Let $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & -1 & 1 \end{pmatrix} \in S^3(\mathbb{R})$. By direct calculation, this matrix has the following eigenvalues and eigenvectors

$$\lambda_1 = 2, \lambda_2 = \lambda_3 = 0, \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Thus, the unrestricted Rayleigh quotient, $f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{x}}$, has the maximum value of $\lambda_1 = 2$, which can be achieved at $\mathbf{x} = \pm \mathbf{v}_1$.

If we now exclude $\mathbf{v}_1$ from the optimization domain, by the preceding theorem, the maximum value of f changes to $\lambda_2 = 0$:

$$\max_{\substack{\mathbf{x}\neq\mathbf{0} \\ \mathbf{v}_1^T\mathbf{x}=0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = 0,$$

which can be attained at $\mathbf{x} = \pm\mathbf{v}_2$.

Note that in this case, because $\lambda_3 = \lambda_2$, the maximum value of the restricted Rayleigh quotient may also be attained at $\mathbf{x} = \pm\mathbf{v}_3$.

In fact, any nonzero vector in the eigenspace corresponding to the repeated eigenvalue 0, $E(0) = \mathrm{span}\{\mathbf{v}_2, \mathbf{v}_3\}$, would maximize the restricted Rayleigh quotient.

The generalized Rayleigh quotients

Def 0.2. For a fixed symmetric matrix $\mathbf{A} \in S^n(\mathbb{R})$ and a positive definite matrix $\mathbf{B} \in S^n_+(\mathbb{R})$ of the same size, a **generalized Rayleigh quotient** corresponding to them is a function $f : \mathbb{R}^n - \{\mathbf{0}\} \longmapsto \mathbb{R}$ defined by

$$f(\mathbf{x}) = \frac{\mathbf{x}^T \mathbf{A} \mathbf{x}}{\mathbf{x}^T \mathbf{B} \mathbf{x}}.$$

Note that if $\mathbf{B} = \mathbf{I}$, then the above problems reduces to an ordinary Rayleigh quotient.

This is also a function defined over $\mathbb{R}^2$ with the origin excluded, and scaling invariant like ordinary Rayleigh quotients:

$$f(k\mathbf{x}) = \frac{(k\mathbf{x})^T\mathbf{A}(k\mathbf{x})}{(k\mathbf{x})^T\mathbf{B}(k\mathbf{x})} = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = f(\mathbf{x}), \qquad \text{for all } \mathbf{x} \neq \mathbf{0}$$

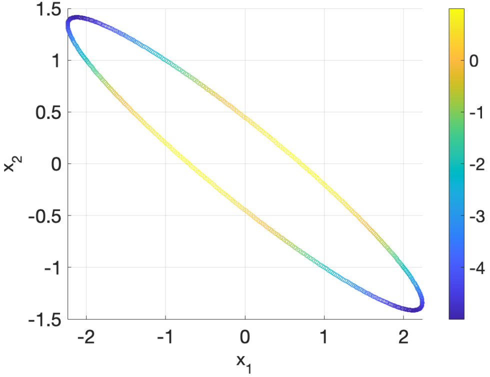
It is essentially a quadratic form $(\mathbf{x}^T\mathbf{A}\mathbf{x})$ over an ellipsoid $(\mathbf{x}^T\mathbf{B}\mathbf{x} = 1)$:

$$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}, \quad \text{for any } \mathbf{x} \text{ satisfying } \mathbf{x}^T\mathbf{B}\mathbf{x} = 1$$

Example 0.4. Given $\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix} \in S^2(\mathbb{R})$ and $\mathbf{B} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix} \in S^2_+(\mathbb{R})$, we have the following generalized Rayleigh quotients:

$$f(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = \frac{2x_1^2 + 2x_2^2 + 6x_1x_2}{2x_1^2 + 5x_2^2 + 6x_1x_2}, \quad \mathbf{x} \neq \mathbf{0} \in \mathbb{R}^2.$$

We plot the values of f (indicated by color) in two dimensions, in order to visualize the function f over the set $\mathbf{x}^T\mathbf{B}\mathbf{x} = 1$.



Theorem 0.4. For any two matrices $\mathbf{A} \in S^n(\mathbb{R})$ and $\mathbf{B} \in S^n_+(\mathbb{R})$, let the largest and smallest generalized eigenvalues of $(\mathbf{A}, \mathbf{B})$ be λ_1 and λ_n, with corresponding generalized eigenvectors $\mathbf{v}_1, \mathbf{v}_n \in \mathbb{R}^n$, respectively. Then the maximum (or minimum) value of the generalized Rayleigh quotient $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ is equal to the largest (or smallest) generalized eigenvalue of $(\mathbf{A}, \mathbf{B})$, achieved by the corresponding generalized eigenvectors:

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = \lambda_1, \quad @\ \mathbf{x} = \pm\mathbf{v}_1$$

$$\min_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = \lambda_n, \quad @\ \mathbf{x} = \pm\mathbf{v}_n$$

Proof. We use the Method of Lagrange multipliers to prove this theorem here. First, the optimization of the generalized Rayleigh quotient

$$\max_{\mathbf{x}\neq\mathbf{0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$$

is equivalent to the following constrained optimization problem:

$$\max_{\mathbf{x}\in\mathbb{R}^n} \mathbf{x}^T\mathbf{A}\mathbf{x} \quad \text{subject to } \mathbf{x}^T\mathbf{B}\mathbf{x} = 1$$

It follows that the Lagrangian function is

$$L(\mathbf{x},\lambda) = \mathbf{x}^T\mathbf{A}\mathbf{x} - \lambda(\mathbf{x}^T\mathbf{B}\mathbf{x} - 1).$$

Now, applying the formulas obtained previously, we have

$$\frac{\partial L}{\partial \mathbf{x}} = 2\mathbf{A}\mathbf{x} - \lambda(2\mathbf{B}\mathbf{x}) = 0 \qquad \longrightarrow \qquad \mathbf{A}\mathbf{x} = \lambda\mathbf{B}\mathbf{x}$$

$$\frac{\partial L}{\partial \lambda} = -(\mathbf{x}^T\mathbf{B}\mathbf{x} - 1) = 0 \qquad \longrightarrow \qquad \mathbf{x}^T\mathbf{B}\mathbf{x} = 1$$

This implies that $\mathbf{x}, \lambda$ must be a (normalized) eigenpair of $\mathbf{A}$. For any solution $\lambda = \lambda_i, \mathbf{x} = \mathbf{v}_i$, the objective function $\mathbf{x}^T\mathbf{A}\mathbf{x}$ takes the value

$$\mathbf{v}_i^T\mathbf{A}\mathbf{v}_i = \mathbf{v}_i^T(\lambda_i\mathbf{B}\mathbf{v}_i) = \lambda_i(\mathbf{v}_i^T\mathbf{B}\mathbf{v}_i) = \lambda_i \cdot 1 = \lambda_i.$$

Therefore, the largest generalized eigenvector $\mathbf{v}_1$ of $(\mathbf{A}, \mathbf{B})$ is a global maximizer, and it yields the absolute maximum value λ_1. Similarly, the smallest generalized eigenvector $\mathbf{v}_n$ is a global minimizer with absolute minimum λ_n.

Remark. **The theorem can also be proved by using linear algebra**: Since $\mathbf{B} \in S_+^n(\mathbb{R})$, it has a square root, $\mathbf{B}^{1/2} \in S_+^n(\mathbb{R})$, which is invertible. Let $\mathbf{y} = \mathbf{B}^{1/2}\mathbf{x}$. Then $\mathbf{x} = \mathbf{B}^{-1/2}\mathbf{y}$. Plug it into the generalized Rayleigh quotient $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ to rewrite it in terms of the new variable $\mathbf{y}$. This will reduce the generalized Rayleigh quotient problem to an ordinary Rayleigh quotient problem, which has already been solved. The rest of the proof is left as homework.

Example 0.5. Consider the two matrices $\mathbf{A} = \begin{pmatrix} 2 & 3 \\ 3 & 2 \end{pmatrix}$ and $\mathbf{B} = \begin{pmatrix} 2 & 3 \\ 3 & 5 \end{pmatrix}$, where $\mathbf{A}$ is symmetric and $\mathbf{B}$ is positive definite. We have already solved the generalized eigenvalue problem $(\mathbf{A}, \mathbf{B})$ previously:

$$\lambda_1 = 1,\ \lambda_2 = -5, \qquad \text{and} \quad \mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \end{pmatrix},\ \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{pmatrix} 3 \\ -2 \end{pmatrix}.$$

Thus, by the preceding theorem, the generalized Rayleigh quotient $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ has a maximum value of $\lambda_1 = 1$ and a minimum value of $\lambda_2 = -5$, attained at the corresponding generalized eigenvectors, $\pm\mathbf{v}_1, \pm\mathbf{v}_2$, respectively.

As for the ordinary Rayleigh quotient, there is a restricted version of the generalized Rayleigh quotient.

Theorem 0.5. Let $\mathbf{A} \in S^n(\mathbb{R})$ and $\mathbf{B} \in S^n_+(\mathbb{R})$ be two fixed matrices with generalized eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$ and eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_n \in \mathbb{R}^n$, that is, $\mathbf{A}\mathbf{v}_i = \lambda_i \mathbf{B}\mathbf{v}_i$ for each $i = 1, \ldots, n$. We have

$$\max_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{v}_1^T \mathbf{B}\mathbf{x} = 0}} \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}} = \lambda_2 \quad (\text{when } \mathbf{x} = \pm \mathbf{v}_2)$$

$$\min_{\substack{\mathbf{x} \neq \mathbf{0} \\ \mathbf{v}_1^T \mathbf{B}\mathbf{x} = \mathbf{v}_2^T \mathbf{B}\mathbf{x} = 0}} \frac{\mathbf{x}^T \mathbf{A}\mathbf{x}}{\mathbf{x}^T \mathbf{B}\mathbf{x}} = \lambda_3 \quad (\text{when } \mathbf{x} = \pm \mathbf{v}_3)$$

and so on.

Example 0.6. Let $\mathbf{A} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 3 & 1 \\ 0 & 1 & 3 \end{pmatrix} \in S^3(\mathbb{R}), \mathbf{B} = \mathrm{diag}(1, 2, 2) \in S^3_+(\mathbb{R})$.

By direct calculation, the generalized eigenvalues and eigenvectors of $(\mathbf{A}, \mathbf{B})$ are

$$\lambda_1 = 2, \lambda_2 = 1, \lambda_3 = 0; \quad \mathbf{v}_1 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \frac{1}{2}\begin{pmatrix} 0 \\ 1 \\ -1 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

Thus, the unrestricted generalized Rayleigh quotient, $f(\mathbf{x}) = \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ over $\mathbb{R}^n - \{\mathbf{0}\}$, has the maximum value of $\lambda_1 = 2$, which can be achieved at $\mathbf{x} = \pm\mathbf{v}_1$.

If we now exclude $\mathbf{v}_1$ from the optimization domain (and consider only the orthogonal complement of it), by the preceding theorem, the maximum value of f changes to $\lambda_2 = 1$:

$$\max_{\substack{\mathbf{x}\neq\mathbf{0}\\ \mathbf{v}_1^T\mathbf{B}\mathbf{x}=0}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}} = 1,$$

which can be attained at $\mathbf{x} = \pm\mathbf{v}_2$.

Applications of Rayleigh quotients

Rayleigh quotients have many applications. Later in this course, we will cover the following:

- **PCA**: $\max_{\mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{\Sigma}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}$ ($\mathbf{\Sigma}$: covariance matrix)
- **LDA**: $\max_{\mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{S}_b\mathbf{v}}{\mathbf{v}^T\mathbf{S}_w\mathbf{v}}$ ($\mathbf{S}_b$: between-class scatter, $\mathbf{S}_w$: within-class scatter)
- **Laplacian Eigenmaps** (and spectral clustering): $\min_{\substack{\mathbf{v}\neq\mathbf{0}\\ \mathbf{v}^T\mathbf{D}\mathbf{1}=0}} \frac{\mathbf{v}^T\mathbf{L}\mathbf{v}}{\mathbf{v}^T\mathbf{D}\mathbf{v}}$ ($\mathbf{L}$: graph Laplacian matrix, $\mathbf{D}$: degree matrix)