San José State University

Math 253: Mathematical Methods for Data Visualization

# Lecture 4: Rayleigh Quotients

Dr. Guangliang Chen

# Outline

- Quadratic forms
- Positive (semi)definite matrices
- Rayleigh quotients

# Recall

... that we have reviewed linear algebra up to **symmetric matrices**, which are square matrices $\mathbf{A}$ satisfying $\mathbf{A}^T = \mathbf{A}$.

Symmetric matrices $\mathbf{A} \in \mathbb{R}^{n\times n}$ have many nice properties:

- All their eigenvalues are real numbers (no complex eigenvalues)
- They are orthogonally diagonalizable, i.e., there exist an orthogonal matrix $\mathbf{Q}$ and a diagonal matrix $\mathbf{\Lambda}$, both of the same size as $\mathbf{A}$, such that

$$\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T \qquad \longleftarrow \text{spectral decomposition of } \mathbf{A}$$

  We also know that $\mathbf{\Lambda}$ consists of the eigenvalues of $\mathbf{A}$ along the diagonal, and $\mathbf{Q}$ has the corresponding orthonormal eigenvectors in columns.

Another use of symmetric matrices is to define the so-called quadratic forms.

**Def 0.1.** Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a symmetric matrix. A quadratic form corresponding to $\mathbf{A}$ is a function $Q : \mathbb{R}^n \mapsto \mathbb{R}$ with

$$Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x}, \qquad \text{for all} \quad \mathbf{x} \in \mathbb{R}^n$$

**Remark**. A quadratic form is a polynomial with terms all of second order:

$$\mathbf{x}^T\mathbf{A}\mathbf{x} = \sum_{i=1}^n \sum_{j=1}^n a_{ij}x_ix_j$$

For example, if $\mathbf{A} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$ and $\mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}$, then

$$Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} = x_1^2 + 2x_2^2 + 6x_1x_2$$

## Positive (semi)definite matrices

**Def 0.2.** A symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$ is said to be **positive semidefinite (PSD)** if the corresponding quadratic form $Q(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$.

If the equality holds true only for $\mathbf{x} = \mathbf{0}$ (i.e., $\mathbf{x}^T\mathbf{A}\mathbf{x} > 0$ for all $\mathbf{x} \neq \mathbf{0}$), then $\mathbf{A}$ is said to be **positive definite (PD)**.

*Theorem* 0.1. A symmetric matrix is positive definite (semidefinite) if and only if all of its eigenvalues are positive (nonnegative).

**Example 0.1.** Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 5 \end{pmatrix}, \mathbf{B} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}, \mathbf{C} = \begin{pmatrix} 1 & 3 \\ 3 & 2 \end{pmatrix}$. According to the theorem, $\mathbf{A}$ is positive definite, $\mathbf{B}$ is positive semidefinite, and $\mathbf{C}$ is neither.

The following result will be needed later.

*Theorem* 0.2. For any rectangular matrix $\mathbf{A} \in \mathbb{R}^{m\times n}$, both $\mathbf{A}\mathbf{A}^T \in \mathbb{R}^{m\times m}$ and $\mathbf{A}^T\mathbf{A} \in \mathbb{R}^{n\times n}$ are square, symmetric, and positive semidefinite.

*Proof.* It is obvious that $\mathbf{A}^T\mathbf{A}$ is square $(n \times n)$ and symmetric:

$$(\mathbf{A}^T\mathbf{A})^T = \mathbf{A}^T(\mathbf{A}^T)^T = \mathbf{A}^T\mathbf{A}$$

To show that it is positive semidefinite, consider the quadratic form: For any $\mathbf{x} \in \mathbb{R}^n$,

$$\mathbf{x}^T(\mathbf{A}^T\mathbf{A})\mathbf{x} = (\mathbf{x}^T\mathbf{A}^T)(\mathbf{A}\mathbf{x}) = (\mathbf{A}\mathbf{x})^T(\mathbf{A}\mathbf{x}) = \|\mathbf{A}\mathbf{x}\|^2 \geq 0.$$

The proof for the other product $\mathbf{A}\mathbf{A}^T$ is similar. □

## Matrix square roots

**Problem**: Let $\mathbf{A} \in \mathbb{R}^{n\times n}$ be a PSD matrix. Find another matrix $\mathbf{B}$ of the size such that $\mathbf{A} = \mathbf{B}^2$. We call $\mathbf{B}$ the square root of $\mathbf{A}$ and denote it by $\mathbf{B} = \mathbf{A}^{1/2}$.

*Solution*. Since $\mathbf{A}$ is symmetric and PSD, there exist an orthogonal matrix $\mathbf{Q} \in \mathbb{R}^{n\times n}$ and a diagonal matrix $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ with all $\lambda_i \geq 0$ such that $\mathbf{A} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^T$.

Define $\mathbf{\Lambda}^{1/2} = \mathrm{diag}(\lambda_1^{1/2}, \ldots, \lambda_n^{1/2})$. Clearly, $\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2} = \mathbf{\Lambda}$.

Let $\mathbf{B} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T$. Then

$$\mathbf{B}^2 = (\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T)(\mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T) = \mathbf{Q}\underbrace{\mathbf{\Lambda}^{1/2}\mathbf{\Lambda}^{1/2}}_{\mathbf{\Lambda}}\mathbf{Q}^T = \mathbf{A}.$$

**Answer**. $\mathbf{B} = \mathbf{A}^{1/2} = \mathbf{Q}\mathbf{\Lambda}^{1/2}\mathbf{Q}^T \quad \longleftarrow$ still a PSD matrix!

**Example 0.2.** Let $\mathbf{A} = \begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix}$, which is PSD because it has two nonnegative eigenvalues $\lambda_1 = 5, \lambda_2 = 0$. To find the matrix square root of $\mathbf{A}$, we need to find its orthogonal diagonalization:

$$\begin{pmatrix} 1 & 2 \\ 2 & 4 \end{pmatrix} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} 5 & \\ & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^T$$

It follows that

$$\mathbf{A}^{1/2} = \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix} \begin{pmatrix} \sqrt{5} & \\ & 0 \end{pmatrix} \begin{pmatrix} \frac{1}{\sqrt{5}} & -\frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{1}{\sqrt{5}} \end{pmatrix}^T = \begin{pmatrix} \frac{1}{\sqrt{5}} & \frac{2}{\sqrt{5}} \\ \frac{2}{\sqrt{5}} & \frac{4}{\sqrt{5}} \end{pmatrix}$$

# Rayleigh quotients

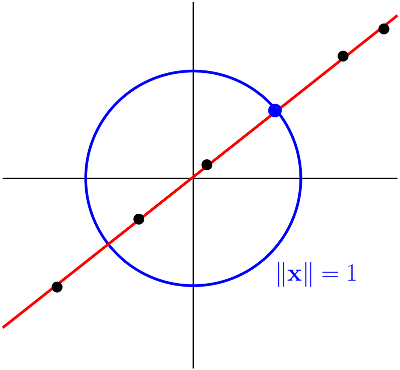
**Problem** (constrained optimization). Given a symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, find the extreme values of the associated quadratic form over the unit sphere in $\mathbb{R}^n$:

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

**Equivalent problem** (unconstrained):

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

(Illustration when $n = 2$)



**Def 0.3.** For a fixed symmetric matrix $\mathbf{A}$, the normalized quadratic form $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$ is called a **Rayleigh quotient**.

Given a positive definite matrix $\mathbf{B}$ of the same size, the quantity $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ is called a **generalized Rayleigh quotient**.

Rayleigh quotients have many applications:

- **PCA**: $\max_{\mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\boldsymbol{\Sigma}\mathbf{v}}{\mathbf{v}^T\mathbf{v}}$ ($\boldsymbol{\Sigma}$: 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 matrix, $\mathbf{S}_w$: within-class scatter matrix)
- **Spectral clustering**: $\max_{\mathbf{v}\neq\mathbf{0}} \frac{\mathbf{v}^T\mathbf{L}\mathbf{v}}{\mathbf{v}^T\mathbf{D}\mathbf{v}}$ ($\mathbf{L}$: graph Laplacian, $\mathbf{D}$: degree matrix)

*Theorem* 0.3. For any given symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$,

$$\max_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{o}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_{\max} \quad \text{(when } \mathbf{x} = \text{“largest” eigenvector of } \mathbf{A}\text{)}$$

$$\min_{\mathbf{x}\in\mathbb{R}^n:\,\mathbf{x}\neq\mathbf{o}} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}} = \lambda_{\min} \quad \text{(when } \mathbf{x} = \text{“smallest” eigenvector of } \mathbf{A}\text{)}$$

**Example 0.3.** For the matrix $\mathbf{A}$ in the preceding example,

- The maximum of the Rayleigh quotient is 5, achieved when $\mathbf{x} = \frac{1}{\sqrt{5}}(1,2)^T$,
- The minimum is 0, achieved when $\mathbf{x} = \frac{1}{\sqrt{5}}(-2,1)^T$

The overall range of the Rayleigh quotient $Q(\mathbf{x}) = \frac{x_1^2+4x_2^2+4x_1x_2}{x_1^2+x_2^2}$ is thus $[0,5]$.

We prove the theorem on the preceding slide in two ways.

(1) **Linear algebra approach**:

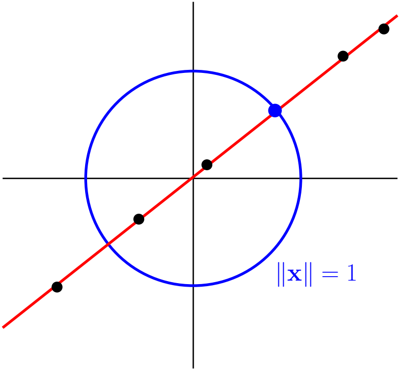
$$\max_{\mathbf{x}\neq\mathbf{0}\in\mathbb{R}^n} \frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{x}}$$

Since the Rayleigh quotient is scaling invariant, we only need to focus on the unit sphere:

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

(2) **Multivariable calculus approach**:

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



## Linear algebra approach

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

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

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

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

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

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

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 objective function attains its maximum value $\mathbf{y}^T \mathbf{\Lambda} \mathbf{y} = \lambda_1$.

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

$$\mathbf{x}^* = \mathbf{Q}\mathbf{y}^* = \mathbf{Q}(\pm \mathbf{e}_1) = \pm \mathbf{q}_1.$$

In conclusion, when $\mathbf{x} = \pm \mathbf{q}_1$ (largest eigenvector), $\mathbf{x}^T \mathbf{A} \mathbf{x}$ attains its maximum value $\lambda_1$ (largest eigenvalue).

## 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 $\frac{\partial L}{\partial \mathbf{x}} = (\frac{\partial L}{\partial x_1}, \ldots, \frac{\partial L}{\partial x_n})^T, \frac{\partial L}{\partial \lambda}$ 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 on next slide.

*Proposition* 0.4. For any fixed symmetric matrix $\mathbf{A} \in \mathbb{R}^{n\times n}$, fixed rectangular matrix $\mathbf{B} \in \mathbb{R}^{m\times n}$ and fixed vector $\mathbf{a} \in \mathbb{R}^n$, we have

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

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

*Proof.* Each of the top two identities can be verified by direct calculation of the $k$th partial derivative, for each $1 \le k \le n$:

$$\frac{\partial}{\partial x_k}\left(\mathbf{a}^T\mathbf{x}\right) = \frac{\partial}{\partial x_k}\left(\sum a_i x_i\right) = a_k$$

$$\frac{\partial}{\partial x_k}\left(\|\mathbf{x}\|^2\right) = \frac{\partial}{\partial x_k}\left(\sum x_i^2\right) = 2x_k.$$

For the third identity involving $\mathbf{x}^T\mathbf{A}\mathbf{x}$,

$$\begin{aligned}
\frac{\partial}{\partial x_k}(\mathbf{x}^T\mathbf{A}\mathbf{x}) &= \frac{\partial}{\partial x_k}\left(\sum_i\sum_j a_{ij}x_ix_j\right) \\
&= \frac{\partial}{\partial x_k}\left(\sum_{\substack{j\neq k\\ i=k}} a_{kj}x_kx_j + \sum_{\substack{i\neq k\\ j=k}} a_{ik}x_ix_k + a_{kk}x_k^2\right) \\
&= \sum_{j\neq k} a_{kj}x_j + \sum_{i\neq k} a_{ik}x_i + 2a_{kk}x_k \\
&= \sum_j a_{kj}x_j + \sum_i x_ia_{ik} \\
&= \mathbf{A}(k,:)\mathbf{x} + \mathbf{x}^T\mathbf{A}(:,k) \\
&= 2\mathbf{A}(k,:)\mathbf{x} \qquad (\text{since } \mathbf{A} \text{ is symmetric})
\end{aligned}$$

Collectively, we have

$$\frac{\partial}{\partial \mathbf{x}}(\mathbf{x}^T\mathbf{A}\mathbf{x}) = \begin{bmatrix} \frac{\partial}{\partial x_1}(\mathbf{x}^T\mathbf{A}\mathbf{x}) \\ \vdots \\ \frac{\partial}{\partial x_n}(\mathbf{x}^T\mathbf{A}\mathbf{x}) \end{bmatrix} = \begin{bmatrix} 2\mathbf{A}(1,:)\mathbf{x} \\ \vdots \\ 2\mathbf{A}(n,:)\mathbf{x} \end{bmatrix} = 2\mathbf{A}\mathbf{x}$$

The last identity can then be verified by writing

$$\|\mathbf{B}\mathbf{x}\|^2 = (\mathbf{B}\mathbf{x})^T(\mathbf{B}\mathbf{x}) = \mathbf{x}^T(\mathbf{B}^T\mathbf{B})\mathbf{x}$$

and applying the third identity. $\square$

Now, applying the formulas obtained previously, we have

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

This implies that $\mathbf{x}, \lambda$ must be an eigenpair of $\mathbf{A}$. For any solution $\lambda = \lambda_i, \mathbf{x} = \mathbf{v}_i$, the objective function 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 $\lambda_1$ of $\mathbf{A}$) is the global maximizer, and it yields the absolute maximum value $\lambda_1$. Similarly, the eigenvector $\mathbf{v}_n$ corresponding to the smallest eigenvalue $\lambda_n$ is the global minimizer with absolute minimum $\lambda_n$.

## The generalized Rayleigh quotient problem

*Corollary* 0.5. For a fixed symmetric matrix $\mathbf{A}$, and a fixed positive definite matrix $\mathbf{B}$ of the same size, the extreme values $\lambda$ of the generalized Rayleigh quotient $\frac{\mathbf{x}^T\mathbf{A}\mathbf{x}}{\mathbf{x}^T\mathbf{B}\mathbf{x}}$ (and the corresponding vectors $\mathbf{v}$) satisfy

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{B}\mathbf{v} \qquad \Longleftrightarrow \qquad \mathbf{B}^{-1}\mathbf{A}\mathbf{v} = \lambda\mathbf{v}$$

**Remark**. The left equation is called a generalized eigenvalue problem, which can be solved easily in MATLAB:

- $E = eig(A, B)$ produces a column vector E containing the generalized eigenvalues of square matrices A and B.
- $[V, D] = eig(A, B)$ produces a diagonal matrix D of generalized eigenvalues and a full matrix V whose columns are the corresponding eigenvectors.

*Proof.* There are two ways to prove this result:

- **Substitution method**: Since $\mathbf{B}$ is PD, it has a square root, denoted as $\mathbf{B}^{1/2}$ (which is also PD and thus invertible). Let $\mathbf{y} = \mathbf{B}^{1/2}\mathbf{x}$. Then the denominator can be written as

$$\mathbf{x}^T\mathbf{B}\mathbf{x} = \mathbf{x}^T\mathbf{B}^{1/2}\mathbf{B}^{1/2}\mathbf{x} = \mathbf{y}^T\mathbf{y}$$

  Substitute $\mathbf{x} = (\mathbf{B}^{1/2})^{-1}\mathbf{y} \overset{\text{denote}}{=} \mathbf{B}^{-1/2}\mathbf{y}$ into the numerator to rewrite it in terms of the new variable $\mathbf{y}$. This will convert the generalized Rayleigh quotient problem back to a regular Rayleigh quotient problem, which has been solved. The rest of the proof is left as homework.

- **Method of Lagrange multipliers**: 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:

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

Now, we can apply the method of Lagrange multipliers with the Lagrangian function

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

The remaining steps are also left as homework.

□