planetmath.org

Math for the people, by the people.

Rayleigh-Ritz theorem

Canonical name	RayleighRitzTheorem
Date of creation	2013-03-22 15:37:18
Last modified on	2013-03-22 15:37:18
Owner	gufotta (12050)
Last modified by	gufotta (12050)
Numerical id	8
Author	gufotta (12050)
Entry type	Theorem
Classification	msc 15A18

Let $A \in \mathbf{C}^{n\times n}$ be a Hermitian matrix. Then its eigenvectors are the critical points (vectors) of the "Rayleigh quotient", which is the real function $R : \mathbb{C}^n \backslash \{\mathbf{0}\} \to \mathbb{R}$

$$R(\mathbf{x}) = \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H \mathbf{x}}, \|\mathbf{x}\| \neq 0$$

and its eigenvalues are its values at such critical points.

As a consequence, we have:

$$\lambda_{max} = \max_{\|\mathbf{x}\|\neq 0} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

and

$$\lambda_{min} = \min_{\|\mathbf{x}\|\neq 0} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

Proof:

First of all, let's observe that for a hermitian matrix, the number $\mathbf{x}^H A\mathbf{x}$ is a real one (actually, $<\mathbf{x}, A\mathbf{x}> = \mathbf{x}^H A\mathbf{x} = (A^H\mathbf{x})^H\mathbf{x} = (A\mathbf{x})^H\mathbf{x} = <A\mathbf{x}, \mathbf{x}> = <\mathbf{x}, A\mathbf{x}>^*$, whence $<\mathbf{x}, A\mathbf{x}> = \mathbf{x}^H A\mathbf{x}$ is real), so that the Rayleigh quotient is real as well.

Let's now compute the critical points $\overline{\mathbf{x}}$ of the Rayleigh quotient, i.e. let's solve the equations system $\frac{dR(\overline{\mathbf{x}})}{d\mathbf{x}} = \mathbf{0}^T$. Let's write $\mathbf{x} = \mathbf{x}^{(R)} + j\mathbf{x}^{(I)}$, $\mathbf{x}^{(R)}$ and $\mathbf{x}^{(I)}$ being respectively the real and imaginary part of $\mathbf{x}$. We have:

$$\frac{dR(\mathbf{x})}{d\mathbf{x}} = \frac{dR(\mathbf{x})}{d\mathbf{x}^{(R)}} + j\frac{dR(\mathbf{x})}{d\mathbf{x}^{(I)}}$$

so that we must have:

$$\frac{dR(\overline{\mathbf{x}})}{d\mathbf{x}^{(R)}} = \frac{dR(\overline{\mathbf{x}})}{d\mathbf{x}^{(I)}} = \mathbf{0}^T$$

Using derivatives rules, we obtain:

$$\frac{dR(\mathbf{x})}{d\mathbf{x}^{(R)}} = \frac{d}{d\mathbf{x}^{(R)}}\left(\frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}\right) = \frac{\frac{d\left(\mathbf{x}^H A\mathbf{x}\right)}{d\mathbf{x}^{(R)}}\mathbf{x}^H\mathbf{x} - \mathbf{x}^H A\mathbf{x}\frac{d\left(\mathbf{x}^H\mathbf{x}\right)}{d\mathbf{x}^{(R)}}}{\left(\mathbf{x}^H\mathbf{x}\right)^2} = \frac{\frac{d\left(\mathbf{x}^H A\mathbf{x}\right)}{d\mathbf{x}^{(R)}} - R(\mathbf{x})\frac{d\left(\mathbf{x}^H\mathbf{x}\right)}{d\mathbf{x}^{(R)}}}{\mathbf{x}^H\mathbf{x}}.$$

Applying matrix calculus rules, we find:

$$\frac{d\left(\mathbf{x}^H A\mathbf{x}\right)}{d\mathbf{x}^{(R)}} = \mathbf{x}^H A\frac{d\mathbf{x}}{d\mathbf{x}^{(R)}} + \mathbf{x}^T A^T\frac{d\mathbf{x}^*}{d\mathbf{x}^{(R)}} = \mathbf{x}^H A + \mathbf{x}^T A^T = \mathbf{x}^H A + \left(\mathbf{x}^H A^H\right)^* =$$

and since $A = A^H$,

$$= \mathbf{x}^H A + \left(\mathbf{x}^H A\right)^* = 2\Re\left(\mathbf{x}^H A\right).$$

In a similar way, we get:

$$\frac{d\left(\mathbf{x}^H \mathbf{x}\right)}{d\mathbf{x}^{(R)}} = 2\Re\left(\mathbf{x}^H\right).$$

Substituting, we obtain:

$$\frac{dR(\mathbf{x})}{d\mathbf{x}^{(R)}} = 2\frac{\Re\left(\mathbf{x}^H A\right) - R(\mathbf{x})\Re\left(\mathbf{x}^H\right)}{\mathbf{x}^H \mathbf{x}}$$

and, after a transposition, equating to the null column vector,

$$\begin{aligned}
\mathbf{0} &= \left(\Re\left(\overline{\mathbf{x}}^H A\right) - R(\overline{\mathbf{x}})\Re\left(\overline{\mathbf{x}}^H\right)\right)^T = \\
&= \Re\left(A^T \overline{\mathbf{x}}^*\right) - R(\overline{\mathbf{x}})\Re\left(\overline{\mathbf{x}}^*\right) = \Re\left((A^H \overline{\mathbf{x}})^*\right) - R(\overline{\mathbf{x}})\Re\left(\overline{\mathbf{x}}^*\right) = \\
&= \Re\left((A\overline{\mathbf{x}})^*\right) - R(\overline{\mathbf{x}})\Re\left(\overline{\mathbf{x}}^*\right) = \Re(A\overline{\mathbf{x}}) - R(\overline{\mathbf{x}})\Re\left(\overline{\mathbf{x}}\right)
\end{aligned}$$

and, since $R(\mathbf{x})$ is real,

$$\Re(A\overline{\mathbf{x}} - R(\overline{\mathbf{x}})\overline{\mathbf{x}}) = \mathbf{0}$$

Let's then evaluate $\frac{dR(\mathbf{x})}{d\mathbf{x}^{(I)}}$:

$$\frac{dR(\mathbf{x})}{d\mathbf{x}^{(I)}} = \frac{d}{d\mathbf{x}^{(I)}}\left(\frac{\mathbf{x}^H A \mathbf{x}}{\mathbf{x}^H \mathbf{x}}\right) = \frac{\frac{d\left(\mathbf{x}^H A \mathbf{x}\right)}{d\mathbf{x}^{(I)}}\mathbf{x}^H \mathbf{x} - \mathbf{x}^H A \mathbf{x}\frac{d\left(\mathbf{x}^H \mathbf{x}\right)}{d\mathbf{x}^{(I)}}}{\left(\mathbf{x}^H \mathbf{x}\right)^2} = \frac{\frac{d\left(\mathbf{x}^H A \mathbf{x}\right)}{d\mathbf{x}^{(I)}} - R(\mathbf{x})\frac{d\left(\mathbf{x}^H \mathbf{x}\right)}{d\mathbf{x}^{(I)}}}{\mathbf{x}^H \mathbf{x}}.$$

Applying again matrix calculus rules, we find:

$$\frac{d\left(\mathbf{x}^H A \mathbf{x}\right)}{d\mathbf{x}^{(I)}} = \mathbf{x}^H A \frac{d\mathbf{x}}{d\mathbf{x}^{(I)}} + \mathbf{x}^T A^T \frac{d\mathbf{x}^*}{d\mathbf{x}^{(I)}} = j\mathbf{x}^H A - j\mathbf{x}^T A^T = j(\mathbf{x}^H A - \left(\mathbf{x}^H A^H\right)^*) =$$

and since $A = A^H$,

$$= j(\mathbf{x}^H A - \left(\mathbf{x}^H A\right)^*) = j(2j\Im\left(\mathbf{x}^H A\right)) = -2\Im\left(\mathbf{x}^H A\right).$$

In a similar way, we get:

$$\frac{d\left(\mathbf{x}^H \mathbf{x}\right)}{d\mathbf{x}^{(I)}} = j\mathbf{x}^H - j\mathbf{x}^T = j(\mathbf{x}^H - \left(\mathbf{x}^H\right)^*) = j(2j\Im(\mathbf{x}^H)) = -2\Im\left(\mathbf{x}^H\right).$$

Substituting, we obtain:

$$\frac{dR(\mathbf{x})}{d\mathbf{x}^{(I)}} = -2\frac{\Im\left(\mathbf{x}^H A\right) - R(\mathbf{x})\Im\left(\mathbf{x}^H\right)}{\mathbf{x}^H\mathbf{x}}$$

and, after a transposition, equating to the null column vector,

$$\begin{aligned}
\mathbf{0} &= \left(\Im\left(\overline{\mathbf{x}}^H A\right) - R(\overline{\mathbf{x}})\Im\left(\overline{\mathbf{x}}^H\right)\right)^T = \\
&= \Im\left(A^T\overline{\mathbf{x}}^*\right) - R(\overline{\mathbf{x}})\Im\left(\overline{\mathbf{x}}^*\right) = \Im\left(\left(A^H\overline{\mathbf{x}}\right)^*\right) - R(\overline{\mathbf{x}})\Im\left(\overline{\mathbf{x}}^*\right) = \\
&= \Im\left((A\overline{\mathbf{x}})^*\right) - R(\overline{\mathbf{x}})\Im\left(\overline{\mathbf{x}}^*\right) = -\Im(A\overline{\mathbf{x}}) + R(\overline{\mathbf{x}})\Im\left(\overline{\mathbf{x}}\right)
\end{aligned}$$

and, since $R(\mathbf{x})$ is real,

$$\Im(A\overline{\mathbf{x}} - R(\overline{\mathbf{x}})\overline{\mathbf{x}}) = \mathbf{0}$$

In conclusion, we have that a stationary vector $\overline{\mathbf{x}}$ for the Rayleigh quotient satisfies the complex eigenvalue equation

$$A\overline{\mathbf{x}} - R(\overline{\mathbf{x}})\overline{\mathbf{x}} = \mathbf{0}$$

whence the thesis.□

Remarks:

1) The two relations $\lambda_{\max} = \max_{\|\mathbf{x}\|\neq 0} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}$ and $\lambda_{\min} = \min_{\|\mathbf{x}\|\neq\mathbf{0}} \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}$ can also be obtained in a simpler way. By Schur's canonical form theorem, any normal (and hence any hermitian) matrix is unitarily diagonalizable, i.e. a unitary matrix U exists such that $A = U\Lambda U^H$ with $\Lambda = diag(\lambda_1, \lambda_2, \cdots, \lambda_n)$. So, since all eigenvalues of a hermitian matrix are real, it's possible to write:

$$\begin{aligned}
\mathbf{x}^H A\mathbf{x} &= \mathbf{x}\ \ {}^H U\Lambda U^H\mathbf{x} = \left(U^H\mathbf{x}\right)^H \Lambda \left(U^H\mathbf{x}\right) = \mathbf{y}^H\Lambda\mathbf{y} = \sum_{i=1}^n \lambda_i \left|y_i\right|^2 \leq \lambda_{\max}\sum_{i=1}^n \left|y_i\right|^2 = \\
&= \lambda_{\max}\mathbf{y}^H\mathbf{y} = \lambda_{\max}\left(U^H\mathbf{x}\right)^H\left(U^H\mathbf{x}\right) = \lambda_{\max}\left(\mathbf{x}^H U U^H\mathbf{x}\right) = \lambda_{\max}\left(\mathbf{x}^H\mathbf{x}\right)
\end{aligned}$$

whence

$$\lambda_{\max} \geq \frac{\mathbf{x}^H A\mathbf{x}}{\mathbf{x}^H\mathbf{x}}$$

But, having defined $A\mathbf{v}_M = \lambda_{\max}\mathbf{v}_M$, we have:

$$\frac{\mathbf{v}_M^H A\mathbf{v}_M}{\mathbf{v}_M^H\mathbf{v}_M} = \frac{\mathbf{v}_M^H\lambda_{\max}\mathbf{v}_M}{\mathbf{v}_M^H\mathbf{v}_M} = \lambda_{\max}$$

so that

$$\lambda_{\max} = \max_{\|\mathbf{x}\| \neq \mathbf{0}} \frac{\mathbf{x}^H A \mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

In a much similar way, we obtain

$$\lambda_{\min} = \min_{\|\mathbf{x}\| \neq \mathbf{0}} \frac{\mathbf{x}^H A \mathbf{x}}{\mathbf{x}^H \mathbf{x}}$$

2) The above relations yield the following noteworthing bounds for the diagonal entries of a hermitian matrix:

$$\lambda_{\min} \leq a_{ii} \leq \lambda_{\max}$$

In fact, having defined

$$\mathbf{e}_i = [\overbrace{0, 0, ..., 0}^{i-1}, 1, 0, ..., 0]^T$$

and observing that $a_{ii} = \frac{\mathbf{e}_i^H A \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{e}_i} = R(\mathbf{e}_i)$, we have:

$$\lambda_{\min} = \min_{\|\mathbf{x}\| \neq \mathbf{0}} R(\mathbf{x}) \leq R(\mathbf{e}_i) \leq \max_{\|\mathbf{x}\| \neq \mathbf{0}} R(\mathbf{x}) = \lambda_{\max}.$$