CS724: Topics in Algorithms
Eigenvalues
Slide Set 7

Prof. Dan A. Simovici
1 Introduction

2 The Characteristic Polynomial of a Matrix

3 Spectra of Hermitian Matrices

4 The Power Method for Eigenvalues

5 MATLAB Computations

6 Diagonalizable Matrices

7 The Rayleigh-Ritz and Related Theorems
The existence of directions that are preserved by linear transformations (which are referred to as eigenvectors) has been discovered by L. Euler in his study of movements of rigid bodies. This work was continued by Lagrange, Cauchy, Fourier and Hermite. The theme of eigenvectors and eigenvalues acquired increasing significance through its applications in heat propagation and stability theory.

Later, Hilbert initiated the study of eigenvalues in functional analysis (in the theory of integral operators). He introduced the terms of “eigenvalue” and “eigenvector”.

The term eigenvalue is a German-English hybrid formed from the German word eigen, which means “own” and the English word “value”.
Let $A \in \mathbb{C}^{n \times n}$ be a square matrix. An eigenvalue of $A$ is a complex number $\lambda$ for which there exists a vector $\mathbf{v}$ such that $A\mathbf{v} = \lambda \mathbf{v}$. The vector $\mathbf{v}$ is an eigenvector and the pair $(\lambda, \mathbf{v})$ is an eigenpair of the matrix $A$.

The set of eigenvalues of a matrix $A$ will be referred to as the spectrum of $A$ and will be denoted by $\text{spec}(A)$.

The invariant subspace of $A \in \mathbb{C}^{n \times n}$ for $\lambda$ coincides with the null space of the matrix $\lambda I_n - A$. 
Example

Let \( A = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix} \). An eigenpair \((\lambda, \mathbf{x})\) must satisfy \( A\mathbf{x} = \lambda\mathbf{x} \), which amounts to

\[
2x_1 + x_2 = \lambda x_1, \\
x_1 + 3x_2 = \lambda x_2.
\]

This is an homogenous system:

\[
(2 - \lambda)x_1 + x_2 = 0, \\
x_1 + (3 - \lambda)x_2 = 0,
\]

which has a non-trivial solution only if

\[
\begin{vmatrix} 2 - \lambda & 1 \\ 1 & 3 - \lambda \end{vmatrix} = 0.
\]
Example

This equality, $\lambda^2 - 5\lambda + 5 = 0$ leads to two eigenvalues

$$\lambda_1 = \frac{1 + \sqrt{5}}{2} \text{ and } \lambda_2 = \frac{1 - \sqrt{5}}{2}.$$ 

For $\lambda_1 = \frac{1 + \sqrt{5}}{2}$ the previous system reduces to

$$-\frac{\sqrt{5} - 1}{2} x_1 + x_2 = 0,$$
$$x_1 + \frac{1 - \sqrt{5}}{2} x_2 = 0,$$

which consists of a single independent equality. Choosing, for example $x_2 = 1$ and $x_1 = \frac{\sqrt{5} - 1}{2}$ we obtain the eigenvector

$$\left( \frac{\sqrt{5} - 1}{2} \right).$$
Theorem

Let $A \in \mathbb{C}^{n \times n}$ be a matrix. The set of eigenvectors of $A$ that correspond to an eigenvalue $\lambda \in \text{spec}(A)$ is a subspace of $\mathbb{C}^n$.

Proof.

Let $x$ and $y$ be two vectors that correspond to an eigenvalue $\lambda$ and let $a, b \in \mathbb{C}$. We have

$$A(ax + by) = aAx + bAy$$
$$= a\lambda x + b\lambda y$$
$$= \lambda(ax + by),$$

hence $ax + by$ is an eigenvector of $A$ that corresponds to the eigenvalue $\lambda$.

The subspace defined above is an invariant subspace of $A$, denoted by $S_{A,\lambda}$. 
If $\lambda$ is an eigenvalue for $A \in \mathbb{C}^{n \times n}$, then $a \lambda$ is an eigenvalue of the matrix $aA$ for every $a \in \mathbb{C}$. Thus, $a \text{spec}(A) = \text{spec}(aA)$.

If $\lambda$ is an eigenvalue of $A \in \mathbb{C}^{n \times n}$, then $Ax = \lambda x$ for some eigenvector $x \in \mathbb{C}^n - \{0\}$. This implies $x^H Ax = \lambda x^H x$, so

$$
\lambda = \frac{x^H Ax}{x^H x}.
$$

In the real case by replacing $x^H$ by $x'$ we obtain $\lambda$ is an eigenvalue and $x$ is an eigenvector that corresponds to $\lambda$, then

$$
\lambda = \frac{x' Ax}{x' x}.
$$
Definition

The geometric multiplicity of an eigenvalue $\lambda$ of a matrix $A \in \mathbb{R}^{n \times n}$ is the number $\text{geomm}(A, \lambda) = \dim(S_{A,\lambda})$.

Equivalently, the geometric multiplicity of $\lambda$ is

$$\text{geomm}(A, \lambda) = \dim(\text{null}(A - \lambda I_n)) = n - \text{rank}(A - \lambda I_n),$$
Theorem

Let $A \in \mathbb{R}^{n \times n}$. We have $0 \in \text{spec}(A)$ if and only if $A$ is a singular matrix. Moreover, in this case, $\text{geom}(A, 0) = n - \text{rank}(A) = \text{dim}(\text{null}(A))$.

Corollary

Let $A \in \mathbb{R}^{n \times n}$. If $0 \in \text{spec}(A)$ and $\text{algm}(A, 0) = 1$, then $\text{rank}(A) = n - 1$.

Proof.

Clearly, we have $\text{geom}(A, 0) = 1$, so $\text{rank}(A) = n - 1$. □
Theorem

Let $A \in \mathbb{C}^{n \times n}$ and let $S \subseteq \mathbb{C}^n$ be an invariant subspace of $A$. If the columns of a matrix $X \in \mathbb{C}^{n \times p}$ form a basis of $S$, then there exists a unique matrix $L \in \mathbb{C}^{p \times p}$ such that $AX = XL$. 
Proof

Let $X = (x_1 \cdots x_p)$. Since $Ax_1 \in S$ it follows that $Ax_1$ can be uniquely expressed as a linear combination of the columns of $X$, that is,

$$Ax_j = x_1 \ell_{1j} + \cdots + x_p \ell_{pj}$$

For $1 \leq i \leq p$ we have

$$Ax_j = X \begin{pmatrix} \ell_{1j} \\ \vdots \\ \ell_{pj} \end{pmatrix}.$$

The matrix $L$ is defined by $L = (l_{ij})$. 
Theorem

Let $A \in \mathbb{C}^{m \times n}$ be a matrix with $\text{rank}(A) = n$ and let $B \in \mathbb{C}^{p \times q}$ be a matrix such that $\text{range}(B) = \text{range}(A) \perp$. Then, $\text{range}(A)$ is an invariant subspace of a matrix $X \in \mathbb{C}^{p \times m}$ if and only if $B^H X A = O_{q,n}$.

Proof.

The following statements are easily seen to be equivalent:
- The subspace $\text{range}(A)$ is an invariant subspace of $X$;
- $X \text{range}(A) \subseteq \text{range}(A)$;
- $X \text{range}(A) \perp \text{range}(A) \perp$;
- $X \text{range}(A) \perp \text{range}(B)$.

The last statement is equivalent to $B^H X A = O_{q,n}$. 
Let $A \in \mathbb{C}^{n \times n}$ be a matrix having the eigenvalues $\lambda_1, \ldots, \lambda_n$. If $x_1, \ldots, x_n$ are $n$ eigenvectors corresponding to these values, then we have

$$Ax_1 = \lambda_1 x_1, \ldots, Ax_n = \lambda_n x_n.$$ 

By introducing the matrix $X = (x_1 \cdots x_n) \in \mathbb{C}^{n \times n}$ these equalities can be written in a concentrated form as

$$AX = X \text{diag}(\lambda_1, \ldots, \lambda_n).$$

Obviously, since the eigenvalues can be listed in several ways, this equality is not unique.
Suppose now that $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are unit vectors and that the eigenvalues $\lambda_1, \ldots, \lambda_n$ are distinct. Then $X$ is a unitary matrix, $X^{-1} = X^H$ and we obtain the equality:

$$A = X \text{diag}(\lambda_1, \ldots, \lambda_n) X^H = \lambda_1 \mathbf{x}_1 \mathbf{x}_1^H + \cdots + \lambda_n \mathbf{x}_n \mathbf{x}_n^H$$

known as the *spectral decomposition* of the matrix $A$. 
If $\lambda$ is an eigenvalue of the matrix $A \in \mathbb{C}^{n \times n}$, there exists a non-zero eigenvector $x \in \mathbb{C}^n$ such that $Ax = \lambda x$. Therefore, the linear system

$$(\lambda I_n - A)x = 0_n$$

has a non-trivial solution. This is possible if and only if $\det(\lambda I_n - A) = 0$, so eigenvalues are the solutions of the equation

$$\det(\lambda I_n - A) = 0.$$ 

Note that $\det(\lambda I_n - A)$ is a polynomial of degree $n$ in $\lambda$, known as the \textit{characteristic polynomial} of the matrix $A$. We denote this polynomial by $p_A$. 
Example

Let

\[
A = \begin{pmatrix}
a_{11} & a_{12} & a_{13} \\
a_{21} & a_{22} & a_{23} \\
a_{31} & a_{32} & a_{33}
\end{pmatrix}
\]

be a matrix in \( \mathbb{C}^{3 \times 3} \). Its characteristic polynomial is

\[
p_A = \begin{vmatrix}
\lambda - a_{11} & -a_{12} & -a_{13} \\
-a_{21} & \lambda - a_{22} & -a_{23} \\
-a_{31} & -a_{32} & \lambda - a_{33}
\end{vmatrix}
= \lambda^3 - (a_{11} + a_{22} + a_{33})\lambda^2
+ (a_{11}a_{22} + a_{22}a_{33} + a_{33}a_{11} - a_{12}a_{21} - a_{23}a_{32} - a_{13}a_{31})\lambda
- (a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{32}a_{21}
- a_{12}a_{21}a_{33} - a_{23}a_{32}a_{11} - a_{13}a_{31}a_{22}).
\]
Theorem

Let $A \in \mathbb{C}^{n \times n}$. Then, $\text{spec}(A) = \text{spec}(A')$ and $\text{spec}(A^H) = \{\overline{\lambda} \mid \lambda \in \text{spec}(A)\}$.

Proof.

We have

$$p_{A'}(\lambda) = \det(\lambda I_n - A') = \det((\lambda I_n - A)') = \det(\lambda I_n - A) = p_A(\lambda).$$

Thus, since $A$ and $A'$ have the same characteristic polynomials, their spectra are the same.

For $A^H$ we can write

$$p_{A^H}(\overline{\lambda}) = \det(\overline{\lambda} I_n - A^H) = \det((\lambda I_n - A)^H) = (p_A(\lambda))^H,$$

which implies the second part of the Theorem.
Equality of spectra of $A$ and $A'$ does not imply that the eigenvectors or the invariant subspaces of the corresponding eigenvalues are identical, as it can be seen from the following example.

**Example**

The matrix $A \in \mathbb{C}^{2 \times 2}$ is defined as

$$A = \begin{pmatrix} a & 0 \\ c & b \end{pmatrix},$$

where $a \neq b$ and $c \neq 0$. It is immediate that $\text{spec}(A) = \text{spec}(A') = \{a, b\}$. For $\lambda_1 = a$ we have the distinct invariant subspaces:

$$S_{A,a} = \left\{ k \begin{pmatrix} a - b \\ c \end{pmatrix} \mid k \in \mathbb{C} \right\}$$

$$S_{A',a} = \left\{ k \begin{pmatrix} 1 \\ 0 \end{pmatrix} \mid k \in \mathbb{C} \right\}.$$
The leading term of the characteristic polynomial of $A$ is generated by $(\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn})$ and equals $\lambda^n$. The fundamental theorem of algebra implies that $p_A$ has $n$ complex roots, not necessarily distinct. Observe also that, if $A$ is a matrix with real entries, the roots are paired as conjugate complex numbers.

**Definition**

The *algebraic multiplicity of an eigenvalue* $\lambda$ of a matrix $A \in \mathbb{C}^{n \times n}$, $\text{algm}(A, \lambda)$ equals $k$ if $\lambda$ is a root of order $k$ of the equation $p_A(\lambda) = 0$. If $\text{algm}(A, \lambda) = 1$, we refer to $\lambda$ as a *simple eigenvalue*.
Example

Let \( A \in \mathbb{R}^{3 \times 3} \) be the matrix

\[
A = \begin{pmatrix}
1 & 1 & 1 \\
0 & 1 & 2 \\
2 & 1 & 0 \\
\end{pmatrix}
\]

The characteristic polynomial of \( A \) is

\[
p_A(\lambda) = \left| \begin{array}{ccc}
\lambda - 1 & -1 & -1 \\
0 & \lambda - 1 & -2 \\
-2 & -1 & \lambda \\
\end{array} \right| = \lambda^3 - 2\lambda^2 - 3\lambda.
\]

Therefore, the eigenvalues of \( A \) are 3, 0 and -1.

The eigenvalues of \( I_3 \) are obtained from the equation

\[
det(\lambda I_3 - I_3) = \left| \begin{array}{ccc}
\lambda - 1 & 0 & 0 \\
0 & \lambda - 1 & 0 \\
0 & 0 & \lambda - 1 \\
\end{array} \right| = (\lambda - 1)^3 = 0.
\]
Example

Let $P(a) \in \mathbb{C}^{n \times n}$ be the matrix

$$P(a) = \begin{pmatrix}
a & 1 & \cdots & 1 \\
1 & a & \cdots & 1 \\
1 & 1 & \cdots & a \\
\vdots & \vdots & \ddots & \vdots \\
1 & 1 & \cdots & a
\end{pmatrix}.$$ 

To find the eigenvalues of $P(a)$ we need to solve the equation

$$\begin{vmatrix}
\lambda - a & -1 & \cdots & -1 \\
-1 & \lambda - a & \cdots & -1 \\
\vdots & \vdots & \ddots & \vdots \\
-1 & -1 & \cdots & \lambda - a
\end{vmatrix} = 0.$$
By adding the first \( n - 1 \) columns to the last and factoring out \( \lambda - (a + n - 1) \), we obtain the equivalent equation

\[
(\lambda - (a + n - 1)) \begin{vmatrix}
\lambda - a & -1 & \cdots & 1 \\
-1 & \lambda - a & \cdots & 1 \\
\vdots & \vdots & \ddots & \vdots \\
-1 & -1 & \cdots & 1 \\
\end{vmatrix} = 0.
\]

Adding the last column from the first \( n - 1 \) columns and expanding the determinant yields the equation

\[
(\lambda - (a + n - 1))(\lambda - a + 1)^{n-1} = 0,
\]

which allows us to conclude that \( P(a) \) has the eigenvalue \( a + n - 1 \) with \( \text{algm}(P(a), a + n - 1) = 1 \) and the eigenvalue \( a - 1 \) with \( \text{algm}(P(a), a - 1) = n - 1 \).
In the special case when $a = 1$ we have $P(1) = J_{n,n}$. Thus, $J_{n,n}$ has the eigenvalue $\lambda_1 = n$ with algebraic multiplicity 1 and the eigenvalue 0 with algebraic multiplicity $n - 1$. 
Definition

A matrix $A \in \mathbb{C}^{n \times n}$ is simple if there exists a linearly independent set of $n$ eigenvectors.

If $A \in \mathbb{C}^{n \times n}$ has $n$ distinct eigenvalues, then $A$ is a simple matrix. The reverse of this statement is false because there exist simple matrices for which not all eigenvalues are distinct. For example, $\text{spec}(I_n) = \{1\}$, but $\{e_1, \ldots, e_n\}$ is a linearly independent set of distinct eigenvectors.
Theorem

Let \( A \in \mathbb{R}^{n \times n} \) be a matrix and let \( \lambda \in \text{spec}(A) \). Then, for any \( k \in \mathbb{P} \), \( \lambda^k \in \text{spec}(A^k) \).

Proof.

The proof is by induction on \( k \geq 1 \). The base step, \( k = 1 \) is immediate. Suppose that \( \lambda^k \in \text{spec}(A^k) \), that is \( A^k \mathbf{x} = \lambda^k \mathbf{x} \) for some \( \mathbf{x} \in V - \{\mathbf{0}\} \). Then, \( A^{k+1} \mathbf{x} = A(A^k \mathbf{x}) = A(\lambda^k \mathbf{x}) = \lambda^k A \mathbf{x} = \lambda^{k+1} \mathbf{x} \), so \( \lambda^{k+1} \in \text{spec}(A^{k+1}) \).
Theorem

Let $A \in \mathbb{R}^{n \times n}$ be a non-singular matrix and let $\lambda \in \text{spec}(A)$. We have $\frac{1}{\lambda} \in \text{spec}(A^{-1})$ and the sets of eigenvectors of $A$ and $A^{-1}$ are equal.

Proof.

Since $\lambda \in \text{spec}(A)$ and $A$ is non-singular we have $\lambda \neq 0$ and $Ax = \lambda x$ for some $x \in V - \{0\}$. Therefore, we have $A^{-1}(Ax) = \lambda A^{-1}x$, which is equivalent to $\lambda^{-1}x = A^{-1}x$, which implies $\frac{1}{\lambda} \in \text{spec}(A^{-1})$. In addition, this implies that the set of eigenvectors of $A$ and $A^{-1}$ are identical.
Theorem

Let $p_A(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n$ be the characteristic polynomial of the matrix $A$. Then, we have $c_i = (-1)^i S_i(A)$ for $1 \leq i \leq n$, where $S_i(A)$ is the sum of all principal minors of order $i$ of $A$. 
Proof

Since \( p_A(\lambda) = \lambda^n + c_1 \lambda^{n-1} + \cdots + c_{n-1} \lambda + c_n \), it is easy to see that the derivatives of \( p_A(\lambda) \) are given by:

\[
\begin{align*}
p^{(1)}_A(\lambda) &= n\lambda^{n-1} + (n-1)c_1 \lambda^{n-2} + \cdots + c_{n-1}, \\
p^{(2)}_A(\lambda) &= n(n-1)\lambda^{n-2} + (n-1)(n-2)c_1 \lambda^{n-3} + \cdots + 2c_{n-2}, \\
\vdots \\
p^{(k)}_A(\lambda) &= n(n-1)\cdots(n-k+1)\lambda^{n-k} + \cdots + k!c_{n-k}, \\
\vdots \\
p^{(n)}_A(\lambda) &= n!c_0.
\end{align*}
\]

This implies

\[
c_{n-k} = k!p^{(k)}_A(0)
\]

for \( 0 \leq k \leq n \).
On other hand, we have

\[ c_{n-k} = \frac{1}{k!} (-1)^k k! S_{n-k}(A) = (-1)^{n-k} S_{n-k}(A), \]

which implies the statement of theorem.
Theorem

All eigenvalues of a Hermitian matrix $A \in \mathbb{C}^{n \times n}$ are real numbers. All eigenvalues of a skew-Hermitian matrix are purely imaginary numbers.

Proof.

Let $(\lambda, x)$ be an eigenpair of the matrix $A$. If $A$ is Hermitian, $x^H A x$ is a real number because it is equal to its conjugate:

$$(x^H A x)^H = x^H A^H x = x^H A x.$$ 

Also, $x^H x$ is a real number for every $x \in \mathbb{C}^n$. Therefore,

$$\lambda = \frac{x^H A x}{x^H x}$$

is a real number.

Suppose now that $B$ is a skew-Hermitian matrix. Then, as above, $x^H A x = -x^H A x$, which implies that the real part of $x^H A x$ is 0. Thus, $x^H A x$ is a purely imaginary number and $\lambda$ is a purely imaginary number.
**Corollary**

*If* $A \in \mathbb{R}^{n \times n}$ *and* $A$ *is a symmetric matrix, then all its eigenvalues are real numbers.*

**Proof.**

Observe that the Hermitian adjoint $A^H$ of a matrix $A \in \mathbb{R}^{n \times n}$ coincides with its transposed matrix $A'$.
Corollary

Let $A \in \mathbb{C}^{m \times n}$ be a matrix. The non-zero eigenvalues of the matrices $AA^H$ and $A^HA$ are positive numbers and they have the same algebraic multiplicities for the matrices $AA^H$ and $A^HA$.

Proof.

We need to verify only that if $\lambda$ is a non-zero eigenvalue of $A^HA$, then $\lambda$ is a positive number. Since $A^HA$ is a Hermitian matrix, $\lambda$ is a real number. The equality $A^HAx = \lambda x$ for some eigenvector $x \neq 0$ implies

$$\lambda \| x \|^2_2 = \lambda x^Hx = (Ax)^H Ax = \| Ax \|^2_2,$$

so $\lambda > 0$. \qed
Corollary

Let $A \in \mathbb{C}^{m \times n}$ be a matrix. The eigenvalues of the matrix $B = A^H A \in \mathbb{C}^{n \times n}$ are real non-negative numbers.

Proof.

The matrix $B$ defined above is clearly Hermitian and, therefore, its eigenvalues are real numbers. Next, if $\lambda$ is an eigenvalue of $B$, then we have:

$$\lambda = \frac{x^H A^H A x}{x^H x} = \frac{(A x)^H A x}{x^H x} = \frac{\| A x \|}{\| x \|} \geq 0,$$

where $x$ is an eigenvector that corresponds to $\lambda$.

Note that if $A$ is a Hermitian matrix, then $A^H A = A^2$, hence the spectrum of $A^H A$ is $\{\lambda^2 \mid \lambda \in \text{spec}(A)\}$. 

\[\square\]
Theorem

If $A \in \mathbb{C}^{n \times n}$ is a Hermitian matrix and $u, v$ are two eigenvectors that correspond to two distinct eigenvalues $\lambda_1$ and $\lambda_2$, then $u \perp v$.

Proof.

We have $Au = \lambda_1 u$ and $Av = \lambda_2 v$. This allows us to write $v^HAu = \lambda_1 v^Hu$. Since $A$ is Hermitian, we have

$$\lambda_1 v^Hu = v^HAu = v^HA^Hu = (Av)^H u = \lambda_2 v^Hu,$$

which implies $v^Hu = 0$, that is, $u \perp v$. \qed
(Ky Fan’s Theorem) Let \( A \in \mathbb{C}^{n \times n} \) be a Hermitian matrix such that \( \text{spec}(A) = \{\lambda_1, \ldots, \lambda_n\} \), where \( \lambda_1 \geq \cdots \geq \lambda_n \) and let \( V = (v^1 \cdots v^n) \) be the matrix whose columns consists of the corresponding unit eigenvectors of \( A \).
Let \( \{x^1, \ldots, x^n\} \) be an orthonormal set of vectors in \( \mathbb{C}^n \). For any positive integer \( q \leq n \), the sums \( \sum_{i=1}^{q} \lambda_i \) and \( \sum_{i=1}^{q} \lambda_{n+1-i} \) are, respectively, the maximum and minimum of \( \sum_{j=1}^{q} x'^j A x^j \). Namely, the maximum is obtained by choosing the vectors \( x^1, \ldots, x^q \) as the first \( q \) columns of \( V \); the minimum is obtained by assigning to \( x^1, \ldots, x^q \) the last \( q \) columns of \( V \).
Proof

Let $x^i = \sum_{p=1}^{n} b_p^i v^p$ the expressions of $x^i$ relative to the basis $V$ for $1 \leq i \leq n$. In matrix form these equalities can be written as:

$$(x^1 \ldots x^n) = (v^1 \ldots v^n) \begin{pmatrix} b_1^1 & \cdots & b_1^n \\ \vdots & \ddots & \vdots \\ b_n^1 & \cdots & b_n^n \end{pmatrix}$$

where $b_p^i = (v^p)'x^i = (x^i)'v_p$. Thus, for $X = (x^1 \ldots x^n)$ and $V = (v^1 \ldots v^n)$ we have $X = VB$, where $B$ is the orthonormal matrix

$$B = \begin{pmatrix} b_1^1 & \cdots & b_1^n \\ \vdots & \ddots & \vdots \\ b_n^1 & \cdots & b_n^n \end{pmatrix}.$$
Proof cont’d

We have:

\[ x^{j'} Ax^j = x^{j'} A b_p^j v^p = b_p^j x^{j'} A v^p \]

\[ = b_p^j (x^j)' \lambda_p v^p = (b_p^j)^2 \lambda_p \]

\[ = \lambda_q \sum_{p=1}^{n} (b_p^j)^2 + \sum_{p=1}^{q} (\lambda_p - \lambda_q) (b_p^j)^2 + \sum_{j=q+1}^{n} (\lambda_j - \lambda_q) (b_p^j)^2. \]

This implies

\[ (x^j)' A x^j \leq \lambda_q + \sum_{p=1}^{q} (\lambda_p - \lambda_q) (b_p^j)^2. \]

Therefore,

\[ \sum_{i=1}^{q} \lambda_i - \sum_{j=1}^{q} (x^j)' A x^j \geq \sum_{i=1}^{q} (\lambda_i - \lambda_q) \left( 1 - \sum_{j=1}^{q} (b_p^j)^2 \right). \]
Again, we have $\sum_{j=1}^{q} (b^j_i)^2 \leq \| x_i \|^2 = 1$, so
$$\sum_{i=1}^{q} (\lambda_i - \lambda_q) \left( 1 - \sum_{j=1}^{q} (b^j_i)^2 \right) \geq 0.$$ The left member of previous inequality becomes 0, when $x^i = v^i$, so
$$\sum_{j=1}^{q} (x^i)' A x^j \leq \sum_{i=1}^{q} \lambda_i.$$ The maximum of $\sum_{j=1}^{q} (x^j)' A x^j$ is obtained when $x^j = v^j$ for $1 \leq j \leq q$, that is, when $X$ consists of the first $q$ columns of $V$ that correspond to eigenvectors of the top $k$ largest eigenvalues.

The argument for the minimum is similar.
An equivalent form of Ky Fan’s Theorem can be obtained by observing that the orthonormality condition of the set \( \{x_1, \ldots, x_q\} \) can be expressed as \( X'X = I_q \), where \( X \in \mathbb{C}^{n \times q} \) is the matrix \( X = (x_1 \cdots x_q) \).

Also, the sum \( \sum_{j=1}^q x_j'Ax_j \) equals \( \text{trace}(X'AX) \).

Thus, Ky Fan’s Theorem is equivalent to the fact that the sums \( \sum_{i=1}^n \lambda_i \) and \( \sum_{i=1}^q \lambda_{n+1-i} \) are, respectively, the maximum and minimum of \( \text{trace}(X'AX) \), where \( X'X = I_q \). In this form, Ky Fan’s Theorem is useful for the discussion of principal component analysis.
Let $A \in \mathbb{R}^{n \times n}$ be a matrix that has eigenvalues $\lambda_1, \ldots, \lambda_n$, such that

$$|\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n|.$$ 

We refer to $\lambda_1$ as the \textit{dominant eigenvalue} of $A$. Observe that a dominant eigenvalues is always real for, otherwise, $A$ would have a pair of eigenvalues $\lambda_1$ and $\bar{\lambda}_1$ having equal absolute values.
To compute the dominant eigenvalue of a matrix and an associated eigenvector we can use a technique known as the *power method*. Let \( \mathbf{v}_1, \ldots, \mathbf{v}_n \) be a sequence of unit eigenvectors corresponding to the eigenvalues \( \lambda_1, \ldots, \lambda_n \) of the matrix \( A \), respectively, where \( \lambda_1 \) is the dominant eigenvalue of \( A \) and \( |\lambda_1| > |\lambda_2| \geq \cdots \geq |\lambda_n| \).
Define a sequence of vectors \( \mathbf{x}_0, \mathbf{x}_1, \ldots \) as follows. The initial vector \( \mathbf{x}_0 \in \mathbb{C}^n \) is any unit vector that is not orthogonal on any eigenvector corresponding to \( \lambda_1 \). Then, \( \mathbf{x}_{k+1} \) is the unit vector given by

\[
\mathbf{x}_{k+1} = \frac{A\mathbf{x}_k}{\|A\mathbf{x}_k\|}
\]

for \( k \in \mathbb{N} \). The unit vector \( \mathbf{x}_k \) can be written as

\[
\mathbf{x}_k = \frac{A^k\mathbf{x}_0}{\|A^k\mathbf{x}_0\|}
\]

for \( k \geq 1 \). Indeed, in the base case \((k = 1)\) this equality clearly holds. Suppose that it holds for \( k \). We have

\[
\mathbf{x}_{k+1} = \frac{A\mathbf{x}_k}{\|A\mathbf{x}_k\|} = \frac{A \frac{A^k\mathbf{x}_0}{\|A^k\mathbf{x}_0\|}}{\|A \frac{A^k\mathbf{x}_0}{\|A^k\mathbf{x}_0\|}\|} = \frac{A^{k+1}\mathbf{x}_0}{\|A^{k+1}\mathbf{x}_0\|}.
\]

We claim that \( \lim_{k \to \infty} \mathbf{x}_k = \mathbf{v}_1 \).
The set \( \{v_1, \ldots, v_n\} \) is linearly independent and this allows us to write

\[
x_0 = a_1v_1 + \cdots + a_nv_n.
\]

By the assumption made concerning \( x_0 \) (as not being orthogonal on any eigenvector corresponding to \( \lambda_1 \)) we have \( a_1 \neq 0 \). Thus,

\[
Ax_0 = a_1Av_1 + \cdots + a_nAv_n = a_1\lambda_1v_1 + \cdots + a_n\lambda_nv_n.
\]

A straightforward induction argument on \( k \geq 1 \) shows that

\[
A^kx_0 = a_1\lambda_1^kv_1 + \cdots + a_n\lambda^nv_n
\]

\[
= a_1\lambda_1^k \left( v_1 + \frac{a_2}{a_1} \left( \frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \frac{a_n}{a_1} \left( \frac{\lambda_n}{\lambda_1} \right)^k v_n \right).
\]
Since $|\lambda_j| < 1$ for $2 \leq j \leq n$, it follows that

$$\lim_{k \to \infty} \frac{A^k x_0}{a_1 \lambda_1^k} = \lim_{k \to \infty} \left( v_1 + \frac{a_2}{a_1} \left( \frac{\lambda_2}{\lambda_1} \right)^k v_2 + \cdots + \frac{a_n}{a_1} \left( \frac{\lambda_n}{\lambda_1} \right)^k v_n \right) = v_1,$$

and the speed of convergence of $\frac{A^k x_0}{a_1 \lambda_1^k}$ to $v_1$ is determined by $\frac{\lambda_2}{\lambda_1}$. This implies

$$\lim_{k \to \infty} \frac{\|A^k x_0\|}{|a_1 \lambda_1^k|} = \|v_1\| = 1.$$

Therefore,

$$\lim_{k \to \infty} x_k = \lim_{k \to \infty} \frac{A^k x_0}{\|A^k x_0\|} = \lim_{k \to \infty} \frac{A^k x_0}{|a_1 \lambda_1^k| \cdot \|A^k x_0\|} = \lim_{k \to \infty} \frac{A^k x_0}{|a_1 \lambda_1^k|} = v_1.$$
Since the limit of the sequence $x_0, \ldots, x_k, \ldots$ is $v_1$ we have a method to approximatively determine a unit vector corresponding to the dominating eigenvalue $\lambda_1$ of $A$. 
Let us apply the iteration method to the symmetric matrix

\[ A = \begin{pmatrix} 3 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 6 \end{pmatrix}, \]

beginning with the vector \( x_0 = 1_3 \). The following MATLAB code computes ten vectors of the sequence \( (x_1, x_2, \ldots) \):

```matlab
x = ones(3,1);
sequence(:,1) = x;
for i=1:10
    x = A*x/norm(A*x);
    sequence(:,i+1)=x;
end
sequence
```

and prints the sequence as

```
1.0000 0.4015 0.3844 0.3772 0.3740 0.3726 0.3719 0.3716 0.3715 0.3714
1.0000 0.5789 0.5678 0.5621 0.5594 0.5581 0.5576 0.5573 0.5571 0.5571
1.0000 0.7097 0.7279 0.7361 0.7398 0.7414 0.7422 0.7425 0.7427 0.7428
```
The power method is limited to the dominant eigenvalue. However, if a close approximative value is known for an eigenvalue \( \lambda_i \), the power method can be applied to compute an eigenvector associated to this eigenvalue. Observe that if \( |\lambda_i - a| < |\lambda_j - a| \) for every \( \lambda_j \in \text{spec}(A) - \{\lambda_i\} \), then \( \frac{1}{\lambda_i - a} \) is the dominant eigenvalue of the matrix \( B = (A - al_n)^{-1} \).
The justification of the algorithm follows from the next statement.

**Theorem**

Let $A \in \mathbb{C}^{n \times n}$ and let $a$ be a number such that $a \notin \text{spec}(A)$. Then, $\mathbf{v}$ is an eigenvector of $A$ that corresponds to the eigenvalue $\lambda$ if and only if $\mathbf{v}$ is an eigenvector of the matrix $B = (A - aI)^{-1}$ that corresponds to the eigenvalue $\frac{1}{\lambda - a} \in \text{spec}(B)$.

**Proof.**

If $\mathbf{v}$ is an eigenvector for $A$ that corresponds to the eigenvalue $\lambda$, then $A\mathbf{v} = \lambda \mathbf{v}$, so $(A - aI)\mathbf{v} = (\lambda - a)\mathbf{v}$. Since $a \notin \text{spec}(A)$, the matrix $A - aI$ is invertible, so $\frac{1}{\lambda - a} \mathbf{v} = (A - aI)^{-1} \mathbf{v}$, which proves that $\mathbf{v}$ is an eigenvector of $B$ that corresponds to the eigenvalue $\frac{1}{\lambda - a}$. The reverse implication is immediate.
The characteristic polynomial of a matrix \( A \) is computed using the function \texttt{charpoly}(A).

**Example**

For the unit matrix \texttt{eye(3)} the characteristic polynomial is computed as follows:

\[
\texttt{>> charpoly(eye(3))}
\]
\[
\texttt{ans =}
\]
\[
\begin{array}{cccc}
1 & -3 & 3 & -1
\end{array}
\]

This means that the polynomial is

\[
p(\lambda) = \lambda^3 - 3\lambda^2 + 3\lambda - 1 = (\lambda - 1)^3.
\]

Thus, \texttt{eye(3)} has a single eigenvalue \( \lambda = 1 \) with algebraic multiplicity 3.
To compute the eigenvalues of a matrix $A \in \mathbb{C}^{n \times n}$ we can use the command `eig(A)` which returns an $n$-dimensional vector containing the eigenvalues of $A$. The variant $[V,D] = eig(A)$ produces a diagonal matrix $D$ of eigenvalues and a full matrix $V \in \mathbb{C}^{n \times n}$ whose columns are the corresponding eigenvectors so that $AV =VD$.

A faster computation of eigenvalues is obtained using the `eigs`. If $A$ is a large, sparse and square matrix, `eigs(A)` returns a vector that consists of the six largest magnitude eigenvalues.
Example

The eigenvalue 1 of the matrix `eye(3)` can be retrieved as

```matlab
>> eig(eye(3))
ans =
   1
   1
   1
```
The function call \([V,D] = \text{eigs}(A)\) returns a diagonal matrix \(D\) that contains six eigenvalues of \(A\) having the largest absolute values and a matrix \(V\) whose columns are the corresponding eigenvectors. If a parameter \(\text{flag}\) is added \([V,D,\text{flag}] = \text{eigs}(A)\) and \(\text{flag}\) is 0 then all the eigenvalues converged; otherwise not all converged. The call \(\text{eigs}(A,k)\) computes the \(k\) largest magnitude eigenvalues. Various options can be set by specifying additional parameters. For instance, if \(A\) is a symmetric matrix, then the call

\([V,D] = \text{eigs}(A,k,\text{’SA’})\)

will compute the \(k\) smallest eigenvalues of \(A\), return a diagonal matrix \(D\) that contains the eigenvalues and a matrix \(V\) that contains the corresponding eigenvectors.
**Diagonalizable matrices** are square matrices that are similar to diagonal matrices.

**Theorem**

A matrix $A \in \mathbb{C}^{n\times n}$ is diagonalizable if and only if there exists a linearly independent set $\{v_1, \ldots, v_n\}$ of $n$ eigenvectors of $A$. 
Proof

Let $A \in \mathbb{C}^{n \times n}$ such that there exists a set $\{\mathbf{v}_1, \ldots, \mathbf{v}_n\}$ of $n$ eigenvectors of $A$ that is linearly independent and let $P$ be the matrix $(\mathbf{v}_1 \; \mathbf{v}_2 \; \cdots \; \mathbf{v}_n)$ that is clearly invertible. We have:

$$P^{-1}AP = P^{-1}(A\mathbf{v}_1 \; A\mathbf{v}_2 \; \cdots \; A\mathbf{v}_n) = P^{-1}(\lambda_1 \mathbf{v}_1 \; \lambda_2 \mathbf{v}_2 \; \cdots \; \lambda_n \mathbf{v}_n)$$

$$= P^{-1}P \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}.$$ 

Therefore, we have $A = PDP^{-1}$, where

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

so $A \sim D.$
Conversely, suppose that $A$ is diagonalizable, so $AP = PD$, where $D$ is a diagonal matrix and $P$ is an invertible matrix, and let $v_1, \ldots, v_n$ be the columns of the matrix $P$. We have $Av_i = d_{ii}v_i$ for $1 \leq i \leq n$, so each $v_i$ is an eigenvector of $A$. Since $P$ is invertible, its columns are linear independent.
Corollary

If $A \in \mathbb{C}^{n \times n}$ is diagonalizable then the columns of any matrix $P$ such that $D = P^{-1}AP$ is a diagonal matrix are eigenvectors of $A$. Furthermore, the diagonal entries of $D$ are the eigenvalues that correspond to the columns of $P$. 
Corollary

If $A \in \mathbb{C}^{n \times n}$ is a matrix such that $\sum \{\text{geom}(A, \lambda) \mid \lambda \in \text{spec}(A)\} = n$, then $A$ is diagonalizable.

Proof.

Suppose that $\text{spec}(A) = \{\lambda_1, \ldots, \lambda_k\}$ and let $B_k = \{v^{k1}, \ldots, v^{kp_k}\}$ be a basis of the invariant spaces $S_{A, \lambda_k}$, where $\sum_{k=1}^{p} p_k = n$. Then $B = \bigcup_{k=1}^{p} B_k$ is a linearly independent set of eigenvectors, so $A$ is diagonalizable.
Corollary

If the eigenvalues of the matrix $A \in \mathbb{C}^{n \times n}$ are distinct, then $A$ is diagonalizable.

Proof.

The set of eigenvectors of $A$ is linearly independent. The statement follows immediately.
Theorem

Let $A \in \mathbb{C}^{n \times n}$ be a block diagonal matrix,

$$A = \begin{pmatrix}
A_{11} & 0 & \cdots & 0 \\
0 & A_{22} & \cdots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \cdots & A_{mm}
\end{pmatrix}.$$ 

A is diagonalizable if and only if every matrix $A_{ii}$ is diagonalizable for $1 \leq i \leq m$. 

Proof

Suppose that $A$ is a block diagonal matrix which is diagonalizable. Furthermore, suppose that $A_{ii} \in \mathbb{C}^{n_i \times n_i}$ and $\sum_{i=1}^{m} n_i = n$. There exists an invertible matrix $P \in \mathbb{C}^{n \times n}$ such that $P^{-1}AP$ is a diagonal matrix $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$. Let $p_1, \ldots, p_n$ be the columns of $P$, which are eigenvectors of $A$. Each vector $p_i$ is divided into $m$ blocks $p_i^j$ with $1 \leq j \leq m$, where $p_i^j \in \mathbb{C}^{n_j}$. Thus, $P$ can be written as

$$
P = \begin{pmatrix}
    p_1^1 & p_2^1 & \cdots & p_n^1 \\
    p_1^2 & p_2^2 & \cdots & p_n^2 \\
    \vdots & \vdots & \ddots & \vdots \\
    p_1^m & p_2^m & \cdots & p_n^m
\end{pmatrix}.
$$
Proof cont’d

The equality $A \mathbf{p}_i = \lambda_i \mathbf{p}_i$ can be expressed as

$$
\begin{pmatrix}
A_{11} & O & \cdots & O \\
O & A_{22} & \cdots & O \\
\vdots & \vdots & \ddots & \vdots \\
O & O & \cdots & A_{mm}
\end{pmatrix}
\begin{pmatrix}
\mathbf{p}_i^1 \\
\mathbf{p}_i^2 \\
\vdots \\
\mathbf{p}_i^m
\end{pmatrix}
= \lambda_i
\begin{pmatrix}
\mathbf{p}_i^1 \\
\mathbf{p}_i^2 \\
\vdots \\
\mathbf{p}_i^m
\end{pmatrix},
$$

which shows that $A_{jj} \mathbf{p}_i^j = \lambda_i \mathbf{p}_i^j$ for $1 \leq j \leq m$. 
Let \( M^j = (p_1^j, p_2^j, \ldots, p_n^j) \in \mathbb{C}^{n_j \times n} \). We claim that \( \text{rank}(M^j) = n_j \). Indeed if \( \text{rank}(M^j) \) were less than \( n_j \), we would have fewer than \( n \) independent rows \( M^j \) for \( 1 \leq j \leq m \). This, however, would imply that the rank of \( P \) is less than \( n \), which contradicts the invertibility of \( P \). Since there are \( n_j \) linearly independent eigenvectors of \( A_{jj} \), it follows that each block \( A_{jj} \) is diagonalizable.
Conversely, suppose that each $A_{jj}$ is diagonalizable, that is, there exists an invertible matrix $Q_j$ such that $Q_j^{-1}A_{jj}Q_j$ is a diagonal matrix. Then, it is immediate to verify that the block diagonal matrix

$$Q = \begin{pmatrix} Q_1 & O & \cdots & O \\ O & Q_2 & \cdots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \cdots & Q_m \end{pmatrix}$$

is invertible and $Q^{-1}AQ$ is a diagonal matrix.
Definition

Two matrices $A, B \in \mathbb{C}^{n \times n}$ are *simultaneously diagonalizable* if there exists a matrix $T \in \mathbb{C}^{n \times n}$ such that $A = TDT^{-1}$ and $B = TET^{-1}$, where $D$ and $E$ are diagonal matrices in $\mathbb{C}^{n \times n}$. 
Lemma

If $C \in \mathbb{C}^{n \times n}$ is a diagonalizable matrix and $T \in \mathbb{C}^{n \times n}$ is an invertible matrix, then $TCT^{-1}$ is also diagonalizable.

Proof.

Since $C$ is diagonalizable, there exists a diagonal matrix $D$ and an invertible matrix $S$ such that $SCS^{-1} = D$, so $C = S^{-1}DS$. Then,

$$TCT^{-1} = TS^{-1}DST^{-1} = (TS^{-1})D(TS^{-1})^{-1}.$$ 

Therefore,

$$(TS^{-1})^{-1}(TCT^{-1})(TS^{-1}) = D,$$

which shows that $TCT^{-1}$ is diagonalizable.
Theorem

Let $A, B \in \mathbb{C}^{n \times n}$ be two diagonalizable matrices. Then $A$ and $B$ are simultaneously diagonalizable if and only if $AB = BA$. 
Proof

Suppose that $A$ and $B$ are simultaneously diagonalizable, so $A = TDT^{-1}$ and $B = TET^{-1}$, where $D$ and $E$ are diagonal matrices. Then $AB = TDT^{-1}TET^{-1} = TDET^{-1}$ and $BA = TET^{-1}TDT^{-1} = TEDT^{-1}$. Since any two diagonal matrices commute we have $AB = BA$.

Conversely, suppose that $AB = BA$. Since $A$ is diagonalizable there exists a matrix $T$ such that

$$TAT^{-1} = \begin{pmatrix}
\lambda_1 I_{k_1} & O_{k_1,k_2} & \cdots & O_{k_1,k_m} \\
O_{k_2,k_1} & \lambda_2 I_{k_2} & \cdots & O_{k_2,k_m} \\
& \vdots & \ddots & \vdots \\
O_{k_m,k_1} & O_{k_m,k_2} & \cdots & \lambda_m I_{k_m}
\end{pmatrix},$$

where $k_i$ is the multiplicity of $\lambda_i$ for $1 \leq i \leq m$. 
It is easy to see that if \( A \) and \( B \) commute, then \( TAT^{-1} \) and \( TBT^{-1} \) also commute. If we write \( TBT^{-1} \) as a block matrix \( TBT^{-1} = (B_{pq}) \), where \( B_{pq} \in \mathbb{C}^{kp \times kq} \) for \( 1 \leq p, q \leq m \), then the fact that the matrices \( TAT^{-1} \) and \( TBT^{-1} \) commute translates into the equality

\[
\begin{pmatrix}
\lambda_1 B_{11} & \lambda_1 B_{12} & \cdots & \lambda_1 B_{1m} \\
\lambda_2 B_{21} & \lambda_2 B_{22} & \cdots & \lambda_2 B_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_m B_{m1} & \lambda_m B_{m1} & \cdots & \lambda_m B_{mm}
\end{pmatrix}
= \begin{pmatrix}
\lambda_1 B_{11} & \lambda_2 B_{12} & \cdots & \lambda_m B_{1m} \\
\lambda_1 B_{21} & \lambda_2 B_{22} & \cdots & \lambda_m B_{2m} \\
\vdots & \vdots & \ddots & \vdots \\
\lambda_1 B_{m1} & \lambda_2 B_{m1} & \cdots & \lambda_m B_{mm}
\end{pmatrix}.
\]
Proof cont’d

Thus, if \( \lambda_i \neq \lambda_j \) we have \( B_{ij} = O_{k_i,k_j} \), which shows that the matrix \( TBT^{-1} \) is a block diagonal matrix

\[
TBT^{-1} = \begin{pmatrix}
B_{11} & O_{k_1,k_2} & \cdots & O_{k_1,k_m} \\
O_{k_2,k_1} & B_{22} & \cdots & O_{k_2,k_m} \\
\vdots & \vdots & \ddots & \vdots \\
O_{k_m,k_1} & O_{k_m,k_2} & \cdots & B_{mm}
\end{pmatrix}.
\]

Since \( B \) is diagonalizable, it follows that \( TBT^{-1} \) is diagonalizable, so each matrix \( B_{jj} \) is diagonalizable. Let \( W_i \) be a matrix such that \( W_i^{-1}B_{ii}W_i \) is diagonal and let \( W \) be the block diagonal matrix

\[
W = \begin{pmatrix}
W_1 & O & \cdots & O \\
O & W_2 & \cdots & O \\
\vdots & \vdots & \ddots & \vdots \\
O & O & \cdots & W_m
\end{pmatrix}.
\]

Then, both \( W^{-1}TAT^{-1}W \) and \( W^{-1}TBT^{-1}W \) are diagonal matrices, and this implies that \( A \) and \( B \) are simultaneously diagonalizable.
Theorem

If \( A, B \in \mathbb{C}^{n \times n} \) and \( A \sim B \), then the two matrices have the same characteristic polynomials and, therefore, \( \text{spec}(A) = \text{spec}(B) \).

Proof.

Since \( A \sim B \), there exists an invertible matrix \( X \) such that \( A = XBX^{-1} \). Then, the characteristic polynomial \( \text{det}(A - \lambda I_n) \) can be rewritten as

\[
\text{det}(A - \lambda I_n) = \text{det}(XBX^{-1} - \lambda XI_nX^{-1}) = \text{det}(X(B - \lambda I_n)X^{-1}) = \text{det}(X) \text{det}(B - \lambda I_n) \text{det}(X^{-1}) = \text{det}(B - \lambda I_n),
\]

which implies \( \text{spec}(A) = \text{spec}(B) \). \( \square \)
Theorem

If $A, B \in \mathbb{C}^{n \times n}$ and $A \sim B$, then $\text{trace}(A) = \text{trace}(B)$.

Proof.

Since the two matrices are similar, they have the same characteristic polynomials, so both $\text{trace}(A)$ and $\text{trace}(B)$ equal $-c_1$, where $c_1$ is the coefficient of $\lambda^{n-1}$ in both $p_A(\lambda)$ and $p_B(\lambda)$. 

$\square$
Theorem

(Schur’s Triangularization Theorem) Let \( A \in \mathbb{C}^{n \times n} \) be a square matrix. There exists a unitary matrix \( U \in \mathbb{C}^{n \times n} \) and an upper-triangular matrix \( T \in \mathbb{C}^{n \times n} \) such that \( A = U^H TU \) and the diagonal elements of \( T \) are the eigenvalues of \( A \). Moreover, each eigenvalue \( \lambda \) occurs in the sequence of diagonal values a number of \( \text{algm}(A, \lambda) \) times.
Proof

The argument is by induction on $n \geq 1$. The base case, $n = 1$, is trivial. So, suppose that the statement is true for matrices in $\mathbb{C}^{(n-1) \times (n-1)}$. Let $\lambda_1 \in \mathbb{C}$ be an eigenvalue of $A$, and let $u$ be an eigenvector that corresponds to this eigenvalue. We have

$$U^H A U = \begin{pmatrix} \lambda_1 & u^H AV \\ O & V^H AV \end{pmatrix},$$

where $U = (u | V)$ is an unitary matrix.
Proof cont’d

By the inductive hypothesis, since $V^H A V \in \mathbb{C}^{(n-1) \times (n-1)}$, there exists a unitary matrix $S \in \mathbb{C}^{(n-1) \times (n-1)}$ such that $V^H A V = S^H W S$, where $W$ is an upper-triangular matrix. Then, we have

$$U^H A U = \begin{pmatrix} \lambda_1 & u^H V S^H W S \\ 0 & S^H W S \end{pmatrix} = \begin{pmatrix} \lambda_1 & O \\ 0 & W \end{pmatrix},$$

which shows that an upper triangular matrix $T$ that is unitarily similar to $A$ can be defined as

$$T = \begin{pmatrix} \lambda_1 & O \\ 0 & W \end{pmatrix}.$$
Since $T \sim_u A$, it follows that the two matrices have the same characteristic polynomials and therefore, the same spectra and algebraic multiplicities for each eigenvalue.
Example

Let $A \in \mathbb{R}^{3 \times 3}$ be the symmetric matrix

\[
A = \begin{pmatrix}
14 & -10 & -2 \\
-10 & -5 & 5 \\
-2 & 5 & 11
\end{pmatrix}
\]

whose characteristic polynomial is:

\[
p_A(\lambda) = \lambda^3 - 20\lambda^2 - 100\lambda + 2000.
\]

The eigenvalues of $A$ are $\lambda_1 = 20$, $\lambda_2 = 10$ and $\lambda_3 = -10$. 
Example

It is easy to see that

\[ \mathbf{v}_1 = \begin{pmatrix} -2 \\ 1 \\ 1 \end{pmatrix}, \mathbf{v}_2 = \begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}, \mathbf{v}_3 = \begin{pmatrix} -2 \\ -5 \\ 1 \end{pmatrix} \]

are eigenvectors that correspond to the eigenvalues \( \lambda_1, \lambda_2 \) and \( \lambda_3 \), respectively. The corresponding unit vectors are

\[ \mathbf{u}_1 = \begin{pmatrix} -\frac{2}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \\ \frac{1}{\sqrt{6}} \end{pmatrix}, \mathbf{u}_2 = \begin{pmatrix} \frac{1}{\sqrt{5}} \\ 0 \\ \frac{2}{\sqrt{5}} \end{pmatrix}, \mathbf{u}_3 = \begin{pmatrix} -\frac{2}{\sqrt{30}} \\ \frac{5}{\sqrt{30}} \\ \frac{1}{\sqrt{30}} \end{pmatrix}. \]

For \( \mathbf{U} = (\mathbf{u}_1 \mathbf{u}_2 \mathbf{u}_3) \) we have

\[ \mathbf{U}' \mathbf{A} \mathbf{U} = \mathbf{U}'(20\mathbf{u}_1 10\mathbf{u}_2 - 10\mathbf{u}_3) = \text{diag}(20, 10, -10). \]
(Rayleigh-Ritz Theorem) Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix and let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues, where $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Define the Rayleigh-Ritz function $ral_A : \mathbb{R}^n - \{0\} \rightarrow \mathbb{R}$ as

$$ral_A(x) = \frac{x^H A x}{x^H x}$$

Then,

$$\lambda_1 \geq ral_A(x) \geq \lambda_n$$

for $x \in \mathbb{C}^n - \{0_n\}$. 
Proof

Since $A$ is Hermitian, there exists a unitary matrix $P$ and a diagonal matrix $T$ such that $A = P^H TP$ and the diagonal elements of $T$ are the eigenvalues of $A$, that is, $T = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. This allows us to write:

$$x^H Ax = x^H P^H TPx = (Px)^H TPx = \sum_{j=1}^{n} \lambda_i |(Px)_i|^2,$$

which implies

$$\lambda_1 \|Px\|^2 \geq x^H Ax \geq \lambda_n \|Px\|^2.$$
Proof cont’d

Since $P$ is unitary we also have

$$\| Px \|^2 = x^H P^H P x = x^H x,$$

which implies

$$\lambda_1 x^H x \geq x^H Ax \geq \lambda_n x^H x,$$

for $x \in \mathbb{C}^n$. 

Corollary

Let \( A \in \mathbb{C}^{n \times n} \) be a Hermitian matrix and let \( \lambda_1, \lambda_2, \ldots, \lambda_n \) be its eigenvalues, where \( \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n \). We have

\[
\begin{align*}
\lambda_1 &= \max\{ x^H A x \mid x^H x = 1 \}, \\
\lambda_n &= \min\{ x^H A x \mid x^H x = 1 \}.
\end{align*}
\]
Proof.

Note that if $\mathbf{x}$ is an eigenvector that corresponds to $\lambda_1$, then $A\mathbf{x} = \lambda_1 \mathbf{x}$, so $\mathbf{x}^H A \mathbf{x} = \lambda_1 \mathbf{x}^H \mathbf{x}$; in particular, if $\mathbf{x}^H \mathbf{x} = 1$ we have $\lambda_1 = \mathbf{x}^H A \mathbf{x}$, so

$$\lambda_1 = \max \{ \mathbf{x}^H A \mathbf{x} \mid \mathbf{x}^H \mathbf{x} = 1 \}.$$

The equality for $\lambda_n$ can be shown in a similar manner.
We discuss next an important result that is a generalization of Rayleigh-Ritz Theorem. Denote by $S_k(L)$ the collection of all subspaces of dimension $k$ of an $\mathbb{F}$-linear space $L$.

**Theorem**

*(Courant-Fisher Theorem)* Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix having the eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$. We have:

$$
\lambda_k = \min_{S \in S_{n-k+1}(\mathbb{C}^n)} \max_{x \in S} \{ x^H A x \mid \| x \|_2 = 1 \},
$$

and

$$
\lambda_k = \max_{S \in S_k(\mathbb{C}^n)} \min_{x \in S} \{ x^H A x \mid \| x \|_2 = 1 \}.
$$
Proof

Since $A$ is Hermitian, there exists a unitary matrix $U$ and a diagonal matrix $D$ such that $A = U^H DU$ and the diagonal elements of $D$ are the eigenvalues of $A$, that is, $D = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

We prove only that

$$
\lambda_k = \min_{S \in S_{n-k+1}(\mathbb{C}^n)} \max_{\|x\|_2 = 1} \{x^H Dx \}.
$$

The proof of the second part of the theorem is entirely similar.

Let $S$ be a subspace of $\mathbb{C}^n$ with $\dim(S) = n - k + 1$. We have $S \cap \langle e_1, \ldots, e_k \rangle \neq \{0_n\}$ because otherwise, the equality $S \cap \langle e_1, \ldots, e_k \rangle = \{0_n\}$ would imply $\dim(S) \leq n - k$. 

Proof cont’d

Define

\[ \tilde{S} = \{ \mathbf{y} \in S \mid \| \mathbf{y} \| = 1 \}, \]

\[ \hat{S} = \{ \mathbf{y} \in \tilde{S} \mid \mathbf{y} \in \langle e_1, \ldots, e_k \rangle \}. \]

Therefore, \( \hat{S} \) consists of vectors of \( \tilde{S} \) having the form

\[
\mathbf{y} = \begin{pmatrix}
  y_1 \\
  \vdots \\
  y_k \\
  0 \\
  \vdots \\
  0
\end{pmatrix}
\]

such that \( \sum_{i=1}^{k} y_i^2 = 1 \). Thus, for all \( \mathbf{y} \in \hat{S} \) we have:

\[
\mathbf{y}^H D \mathbf{y} = \sum_{i=1}^{k} \lambda_i |y_i|^2 \geq \lambda_k \sum_{i=1}^{k} |y_i|^2 = \lambda_k
\]
Proof cont’d

Since $\hat{S} \subseteq \tilde{S}$ it follows that $\max_{y \in \hat{S}} y^H D y \geq \max_{y \in \tilde{S}} y^H D y \geq \lambda_k$, so

$$\min_{\dim(S) = n-k+1} \max_{x \in S \text{ and } \|x\|_2 = 1} \{x^H D x\} \geq \lambda_k.$$ 

Let now $S$ be $S = \langle e_1, \ldots, e_{k-1} \rangle^\perp$. Clearly, $\dim(S) = n - k + 1$. A vector $y \in S$ has the form

$$y = \begin{pmatrix} 0 \\ \vdots \\ 0 \\ y_k \\ \cdots \\ y_n \end{pmatrix}$$
Proof cont’d

Therefore,

\[ y^H D y = \sum_{i=k}^{n} \lambda_i |y_i|^2 \leq \lambda_i \sum_{i=k}^{n} |y_i|^2 = \lambda_i \]

for all \( y \in \{y \in S \mid \|y\|_2 = 1\} \). This implies

\[ \min_{\text{dim}(S)=n-k+1} \max_{x} \{ x^H D x \mid x \in S \text{ and } \|x\|_2 = 1 \} \leq \lambda_k, \]

which yields the desired equality.
Proof cont’d

The matrices $A$ and $D$ have the same eigenvalues. Also
\[ x^H A x = x^H A x = x^H U^H D U x = (U x)^H D (U x) \]
and \[ \| U x \|_2 = \| x \|_2, \]
because $U$ is a unitary matrix. This yields the first equality of the theorem.
Theorem

(Interlacing Theorem) Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix and let $B = A \begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix}$ be a principal submatrix of $A$, $B \in \mathbb{C}^{k \times k}$. If $\text{spec}(A) = \{\lambda_1, \ldots, \lambda_n\}$ and $\text{spec}(B) = \{\mu_1, \ldots, \mu_k\}$, where $\lambda_1 \geq \cdots \geq \lambda_n$ and $\mu_1 \geq \cdots \geq \mu_k$, then $\lambda_j \geq \mu_j \geq \lambda_{n-k+j}$ for $1 \leq j \leq k$. 
Proof

Let \( \{j_1, \ldots, j_q\} = \{1, \ldots, n\} - \{i_1, \ldots, i_k\} \), where \( j_1 < \cdots < j_q \) and \( k + q = n \). By Courant-Fisher Theorem we have

\[
\lambda_j = \min_W \max_x \{ x^H A x \mid \|x\|_2 = 1 \text{ and } x \in \langle W \rangle^\perp \},
\]

where \( W \) ranges over sets of non-zero vectors in \( \mathbb{C}^n \) containing \( j - 1 \) vectors. Therefore,

\[
\lambda_j \geq \min_W \max_x \{ x^H A x \mid \|x\|_2 = 1 \text{ and } x \in \langle W \rangle^\perp \\
\text{and } x \in \langle e_{j_1}, \ldots, e_{j_q} \rangle^\perp \}
\]

\[
= \min_U \max_y \{ y^H B y \mid \|y\|_2 = 1 \text{ and } y \in \langle U \rangle^\perp = \mu_j, \}
\]

where \( U \) ranges over sets of non-zero vectors in \( \mathbb{C}^k \) containing \( j - 1 \) vectors.
Proof cont’d

Again, by Courant-Fisher Theorem,

\[ \lambda_{n-k+j} = \max_Z \min_x \{ x^H A x \mid \| x \|_2 = 1 \text{ and } x \in \langle Z \rangle^\perp \} \]

where \( Z \) ranges over sets containing \( k - j \) non-zero vectors in \( \mathbb{C}^n \).

Consequently,

\[
\lambda_{n-k+j} \leq \max_Z \min_x \{ x^H A x \mid \| x \|_2 = 1 \text{ and } x \in \langle Z \rangle^\perp \text{ and } x \in \langle e_{j_1}, \ldots, e_{j_q} \rangle^\perp \} \\
= \max_S \min_y \{ y^H B y \mid \| y \|_2 = 1 \text{ and } y \in \langle S \rangle^\perp \} = \mu_j,
\]

where \( S \) ranges over the sets of non-zero vectors in \( \mathbb{C}^k \) containing \( n - j \) vectors.
Corollary

Let $A \in \mathbb{C}^{n \times n}$ be a Hermitian matrix and let $B = A \begin{bmatrix} i_1 & \cdots & i_k \\ i_1 & \cdots & i_k \end{bmatrix}$ be a principal submatrix of $A$, $B \in \mathbb{C}^{k \times k}$. The set $\text{spec}(B)$ contains no more positive eigenvalues than the number of positive eigenvalues of $A$ and no more negative eigenvalues than the number of negative eigenvalues of $A$. 
By Viéte’s Theorem from algebra we have:

\[ \lambda_1 + \cdots + \lambda_n = a_{11} + a_{22} + \cdots + a_{nn} = \text{trace}(A) = -c_1. \]

Another interesting fact is:

\[ \lambda_1 \cdots \lambda_n = \det(A). \]
Theorem

Let $A \in \mathbb{C}^{m\times n}$ and $B \in \mathbb{C}^{n\times m}$ be two matrices. Then the set of non-zero eigenvalues of the matrices $AB \in \mathbb{C}^{m\times m}$ and $BA \in \mathbb{C}^{n\times n}$ are the same and $\text{algm}(AB, \lambda) = \text{algm}(BA, \lambda)$ for each such eigenvalue.
Proof

Consider the following straightforward equalities:

\[
\begin{pmatrix}
I_m & -A \\
O_{n,m} & \lambda I_n
\end{pmatrix}
\begin{pmatrix}
\lambda I_m & A \\
B & I_n
\end{pmatrix}
= \begin{pmatrix}
\lambda I_m - AB & O_{m,n} \\
-\lambda B & \lambda I_n
\end{pmatrix}
\]

\[
\begin{pmatrix}
-I_m & O_{m,n} \\
-B & \lambda I_n
\end{pmatrix}
\begin{pmatrix}
\lambda I_m & A \\
B & I_n
\end{pmatrix}
= \begin{pmatrix}
-\lambda I_m & -A \\
O_{n,m} & \lambda I_n - BA
\end{pmatrix}.
\]
Proof cont’d

Observe that

$$\det \left( \begin{pmatrix} I_m & -A \\ O_{n,m} & \lambda I_n \end{pmatrix} \begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} \right) = \det \left( \begin{pmatrix} -I_m & O_{m,n} \\ -B & \lambda I_n \end{pmatrix} \begin{pmatrix} \lambda I_m & A \\ B & I_n \end{pmatrix} \right),$$

and therefore,

$$\det \begin{pmatrix} \lambda I_m - AB & O_{m,n} \\ -\lambda B & \lambda I_n \end{pmatrix} = \det \begin{pmatrix} -\lambda I_m & -A \\ O_{n,m} & \lambda I_n - BA \end{pmatrix}.$$
Proof cont’d

The last equality amounts to

\[ \lambda^n p_{AB}(\lambda) = \lambda^m p_{BA}(\lambda). \]

Thus, for \( \lambda \neq 0 \) we have \( p_{AB}(\lambda) = p_{BA}(\lambda) \), which gives the desired conclusion.
Corollary

Let

\[ a = \begin{pmatrix} a_1 \\ \vdots \\ a_n \end{pmatrix} \]

be a vector in \( \mathbb{C}^n - \{0\} \). Then, the matrix \( aa^H \in \mathbb{C}^{n \times n} \) has one eigenvalue distinct from 0, and this eigenvalue is equal to \( \| a \|^2 \).

Proof.

The matrix \( aa^H \) has the same non-zero eigenvalues as the matrix \( a^H a \in \mathbb{C}^{1 \times 1} \) and the single eigenvalue of \( a^H a \) is \( a^H a = \| a \|^2 \). \( \square \)