Lecture Notes: Matrix Algebra
Part F: Eigenvectors and Diagonalization

Peter J. Hammond

minor revision 2023 September 15th;
typeset from `matrixAlgF22.tex`

Outline

Complex Eigenvalues

The eigenvalues we are about to define may be complex numbers.

Instead of $\mathbb{R}^n$ we need to consider the linear space $\mathbb{C}^n$
whose elements are *n*-vectors with complex coordinates.

That is, we consider a linear space whose field of scalars is,
instead of the line $\mathbb{R}$ of real numbers a,
the plane $\mathbb{C}$ of complex numbers $a + b\,\mathrm{i}$, where $a, b \in \mathbb{R}$
and i denotes the basic imaginary number that satisfies $\mathrm{i}^2 = -1$.

Nevertheless, we consider here only $n \times n$ matrices $\mathbf{A}$
whose n^2 elements are all real.

Definition

The complex scalar $\lambda \in \mathbb{C}$ is an eigenvalue
just in case the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ has a non-zero solution;
then that solution $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ is an eigenvector.

Complex Eigenvalues Go with Complex Eigenvectors

Proposition

Given any real $n \times n$ matrix $\mathbf{A}$,
suppose that $\lambda \in \mathbb{C} \setminus \mathbb{R}$ is a complex eigenvalue.
Then any eigenvector of $\mathbf{A}$ satisfies $\mathbf{x} \in \mathbb{C}^n \setminus \mathbb{R}^n$.

Proof.

Suppose that the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ has a solution
with $\lambda = \alpha + \mathrm{i}\,\beta \in \mathbb{C}$ and $\mathbf{x} = \mathbf{y} + \mathrm{i}\,\mathbf{z} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.

Writing out $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ in full gives $\mathbf{A}(\mathbf{y} + \mathrm{i}\,\mathbf{z}) = (\alpha + \mathrm{i}\,\beta)(\mathbf{y} + \mathrm{i}\,\mathbf{z})$.

Equating the imaginary parts gives $\mathbf{A}\mathbf{z} = \beta\mathbf{y} + \alpha\mathbf{z}$.

Now, if $\mathbf{z} = \mathbf{0}$, then $\beta\,\mathbf{y} = \mathbf{0}$ but $\mathbf{y} + \mathrm{i}\,\mathbf{z} = \mathbf{y} \neq \mathbf{0}$, so $\beta = 0$.

On the other hand, if $\lambda = \alpha + \mathrm{i}\,\beta \in \mathbb{C} \setminus \mathbb{R}$ is an eigenvalue of $\mathbf{A}$
with eigenvector $\mathbf{x} = \mathbf{y} + \mathrm{i}\,\mathbf{z}$, then $\mathbf{z} \neq \mathbf{0}$ and so $\mathbf{x} \in \mathbb{C}^n \setminus \mathbb{R}^n$. □

So we have proved that real eigenvalues go with real eigenvectors,
whereas complex eigenvalues go with complex eigenvectors.

The Spectrum of a General Real Square Matrix

Definition

Consider any $n \times n$ matrix $\mathbf{A}$ whose elements are all real.

The scalar $\lambda \in \mathbb{C}$ is an eigenvalue of $\mathbf{A}$ just in case the equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ has a non-zero solution $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$.

In this case the solution $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ is an eigenvector, and the pair $(\lambda, \mathbf{x})$ is an eigenpair.

The spectrum of the matrix $\mathbf{A}$ is the set $S_{\mathbf{A}}$ of its eigenvalues.

Let $S_{\mathbf{A}}^{\mathbb{R}} \subset \mathbb{R}$ denote the subset of its real eigenvalues.

Let $S_{\mathbf{A}}^{\mathbb{C}} \subset \mathbb{C} \setminus \mathbb{R}$ denote the subset of its complex eigenvalues, which satisfies $S_{\mathbf{A}}^{\mathbb{C}} = S_{\mathbf{A}} \setminus S_{\mathbf{A}}^{\mathbb{R}}$.

Summary of Main Properties

We will be demonstrating the following properties:

1. $S_{\mathbf{A}}^{\mathbb{R}} \subseteq S_{\mathbf{A}}$ and $\# S_{\mathbf{A}} \leq n$
2. The number $\# S_{\mathbf{A}}^{\mathbb{C}}$ of complex eigenvalues is even, and the members of $S_{\mathbf{A}}^{\mathbb{C}}$ are complex conjugate pairs $\lambda \pm \mu i$ where $\mu \neq 0$.
3. $S_{\mathbf{A}}^{\mathbb{R}} = \emptyset$ is possible in case n is even, but not if n is odd.
4. In case $\mathbf{A}$ is symmetric, one has $S_{\mathbf{A}}^{\mathbb{C}} = \emptyset$ and $S_{\mathbf{A}}^{\mathbb{R}} = S_{\mathbf{A}}$.

The Eigenspace

Given any eigenvalue $\lambda \in \mathbb{C}$,
let $E_\lambda := \{\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\} \mid \mathbf{A}\mathbf{x} = \lambda \mathbf{x}\}$
denote the associated set of eigenvectors.

Given any two eigenvectors $\mathbf{x}, \mathbf{y} \in E_\lambda$
and any two scalars $\alpha, \beta \in \mathbb{C}$, note that

$$\mathbf{A}(\alpha \mathbf{x} + \beta \mathbf{y}) = \alpha \mathbf{A}\mathbf{x} + \beta \mathbf{A}\mathbf{y} = \alpha \lambda \mathbf{x} + \beta \lambda \mathbf{y} = \lambda(\alpha \mathbf{x} + \beta \mathbf{y})$$

Hence the linear combination $\alpha \mathbf{x} + \beta \mathbf{y}$,
unless it is $\mathbf{0}$, is also an eigenvector in E_λ.

It follows that the set $E_\lambda \cup \{\mathbf{0}\}$ is a linear subspace of $\mathbb{C}^n$
which we call the eigenspace associated with the eigenvalue λ.

Powers of a Matrix

Theorem

Suppose that $(\lambda, \mathbf{x})$ *is an eigenpair of the* $n \times n$ *matrix* $\mathbf{A}$*.*

Then $\mathbf{A}^m \mathbf{x} = \lambda^m \mathbf{x}$ *for all* $m \in \mathbb{N}$*.*

Proof.

By definition, $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$.

Premultiplying each side of this equation by the matrix $\mathbf{A}$ gives

$$\mathbf{A}^2 \mathbf{x} = \mathbf{A}(\mathbf{A}\mathbf{x}) = \mathbf{A}(\lambda \mathbf{x}) = \lambda(\mathbf{A}\mathbf{x}) = \lambda(\lambda \mathbf{x}) = \lambda^2 \mathbf{x}$$

As the induction hypothesis,
suppose that $\mathbf{A}^{m-1}\mathbf{x} = \lambda^{m-1}\mathbf{x}$ for any $m = 2, 3, \ldots$

Premultiplying each side of this last equation by the matrix $\mathbf{A}$ gives

$$\mathbf{A}^m \mathbf{x} = \mathbf{A}(\mathbf{A}^{m-1}\mathbf{x}) = \mathbf{A}(\lambda^{m-1}\mathbf{x}) = \lambda^{m-1}(\mathbf{A}\mathbf{x}) = \lambda^{m-1}(\lambda \mathbf{x}) = \lambda^m \mathbf{x}$$

This completes the proof by induction on m. □

Outline

The Characteristic Equation

The equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ holds for $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}_n\}$
if and only if $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}_n\}$ solves $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.

This holds if and only if the matrix $\mathbf{A} - \lambda\mathbf{I}$ is singular,
which holds if and only if λ is a characteristic root
— i.e., it solves the characteristic equation $|\mathbf{A} - \lambda\mathbf{I}| = 0$.

Equivalently, λ is a zero
of the polynomial $P(\lambda) \equiv |\mathbf{A} - \lambda\mathbf{I}|$ of degree n.

Suppose $|\mathbf{A} - \lambda\mathbf{I}| = 0$ has k distinct roots $\lambda_1, \lambda_2, \ldots, \lambda_k$ in $\mathbb{C}$
whose multiplicities are respectively $m_1, m_2, \ldots, m_k$.

This means that

$$\begin{aligned} |\mathbf{A} - \lambda\mathbf{I}| &= (-1)^n(\lambda - \lambda_1)^{m_1}(\lambda - \lambda_2)^{m_2} \cdots (\lambda - \lambda_k)^{m_k} \\ &= (-1)^n \prod_{j=1}^{k}(\lambda - \lambda_j)^{m_j} \end{aligned}$$

The polynomial has degree $m_1 + m_2 + \ldots + m_k$, which equals n.

This implies that $k \leq n$,
so there can be at most n distinct eigenvalues.

Eigenvalues of a 2×2 matrix

Consider the 2×2 matrix $\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$.

The characteristic equation for its eigenvalues is

$$|\mathbf{A} - \lambda \mathbf{I}| = \begin{vmatrix} a_{11} - \lambda & a_{12} \\ a_{21} & a_{22} - \lambda \end{vmatrix} = 0$$

Evaluating the determinant gives the equation

$$\begin{aligned} 0 &= (a_{11} - \lambda)(a_{22} - \lambda) - a_{12}a_{21} \\ &= \lambda^2 - (a_{11} + a_{22})\lambda + (a_{11}a_{22} - a_{12}a_{21}) \\ &= \lambda^2 - (\operatorname{tr} \mathbf{A})\lambda + |\mathbf{A}| \ = \ (\lambda - \lambda_1)(\lambda - \lambda_2) \end{aligned}$$

where the two roots λ_1 and λ_2 of the quadratic equation have:

- a sum $\lambda_1 + \lambda_2$ equal to the trace $\operatorname{tr} \mathbf{A}$ of $\mathbf{A}$ (the sum of its diagonal elements);
- a product $\lambda_1 \cdot \lambda_2$ equal to the determinant $|\mathbf{A}|$ of $\mathbf{A}$.

Let $\mathbf{\Lambda}$ denote the diagonal matrix $\mathbf{diag}(\lambda_1, \lambda_2)$ whose diagonal elements are the eigenvalues.

Note that $\operatorname{tr} \mathbf{A} = \operatorname{tr} \mathbf{\Lambda}$ and $|\mathbf{A}| = |\mathbf{\Lambda}|$.

The Case of a Diagonal Matrix, I

For the diagonal matrix $\mathbf{D} = \mathbf{diag}(d_1, d_2, \ldots, d_n)$
one has $\mathbf{D} - \lambda \mathbf{I} = \mathbf{diag}(d_1 - \lambda, d_2 - \lambda, \ldots, d_n - \lambda)$.

Then the characteristic equation $|\mathbf{D} - \lambda \mathbf{I}| = 0$
takes the degenerate form $\prod_{k=1}^{n} (d_k - \lambda) = 0$.

So the spectrum $S_{\mathbf{D}}$ of $\mathbf{D}$ equals
the subset of the set $\{d_1, d_2, \ldots, d_n\}$ of diagonal elements
that results from dropping repeated elements.

The natural number $\# S_{\mathbf{D}}$ could be anywhere between 1 and n.

The ith component of the vector equation $\mathbf{D}\mathbf{x} = d_k \mathbf{x}$
takes the form $d_i x_i = d_k x_i$,
which has a non-trivial solution if and only if $d_i = d_k$.

For each $k \in \mathbb{N}_n$, the kth column vector $\mathbf{e}^k = (\delta_{jk})_{j=1}^{n}$
of the canonical orthonormal basis of $\mathbb{R}^n$
satisfies the equation $\mathbf{D}\mathbf{e}^k = d_k \mathbf{e}^k$.

Hence $\mathbf{e}^k$ is an eigenvector associated with the eigenvalue d_k.

The Case of a Diagonal Matrix, II

Apart from non-zero multiples of the canonical basis vector $\mathbf{e}^k$, there are other eigenvectors associated with d_k only if a different element d_i of the diagonal also equals d_k.

In fact, the eigenspace associated with each eigenvalue d_k equals the space spanned by the set $\{\mathbf{e}^i \mid d_i = d_k\}$ of canonical basis vectors.

Example

In case $\mathbf{D} = \mathbf{diag}(1, 1, 0)$ the spectrum is $\{0, 1\}$ with:

- the one-dimensional eigenspace

$$E_0 = \{x_3 \, (0, 0, 1)^\top \mid x_3 \in \mathbb{R}\}$$

- the two-dimensional eigenspace

$$E_1 = \{x_1 \, (1, 0, 0)^\top + x_2 \, (0, 1, 0)^\top \mid (x_1, x_2) \in \mathbb{R}^2\}$$

Characterizing 2×2 Orthogonal Matrices

By definition, an orthogonal matrix $\mathbf{P}$ satisfies $\mathbf{P}^\top \mathbf{P} = \mathbf{P}\mathbf{P}^\top = \mathbf{I}$.

In the 2×2 case when $\mathbf{P} = (p_{ij})_{2\times 2}$, the matrix $\mathbf{P}\mathbf{P}^\top$ equals

$$\begin{pmatrix} p_{11} & p_{12} \\ p_{21} & p_{22} \end{pmatrix} \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix} = \begin{pmatrix} (p_{11})^2 + (p_{12})^2 & p_{11}p_{21} + p_{12}p_{22} \\ p_{21}p_{11} + p_{22}p_{12} & (p_{21})^2 + (p_{22})^2 \end{pmatrix}$$

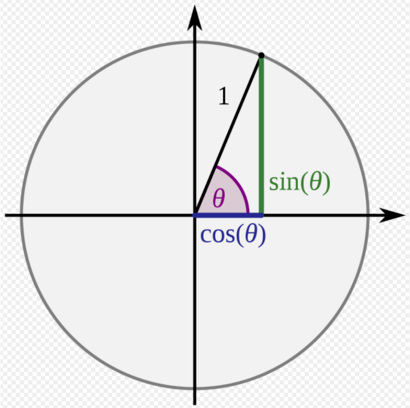
This equals $\mathbf{I}$ if and only if $(p_{11})^2 + (p_{12})^2 = (p_{21})^2 + (p_{22})^2 = 1$ and also $p_{11}p_{21} + p_{12}p_{22} = p_{21}p_{11} + p_{22}p_{12} = 0$.

Inspired by the trigonometric identity $\sin^2 \omega + \cos^2 \omega \equiv 1$,
suppose we put $p_{11} = \cos\theta$ and $p_{22} = \cos\eta$,
along with $p_{12} = -\sin\theta$ and $p_{21} = \sin\eta$.

These choices of sign make $\mathbf{P}$ equal
to the rotation matrix $\mathbf{R}_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
where $\theta \in \mathbb{R}$ is the angle of rotation measured in radians.

This is illustrated in the next slide.

Rotation Matrices Illustrated in an Argand Diagram



Illustrating $P_\theta = \mathbf{R}_\theta \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 1 \\ 0 \end{pmatrix} = \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}$.

Also $P_{\theta+\frac{1}{2}\pi} = \mathbf{R}_\theta \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}$.

Confirming that 2×2 Rotation Matrices are Orthogonal

Given any 2-dimensional rotation matrix $\mathbf{R}_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$, note that

$$\begin{aligned} \mathbf{R}_\theta \mathbf{R}_\theta^\top &= \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{pmatrix} \\ &= \begin{pmatrix} \cos^2\theta + \sin^2\theta & 0 \\ 0 & \sin^2\theta + \cos^2\theta \end{pmatrix} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \end{aligned}$$

Hence $\mathbf{R}_\theta$ is an orthogonal matrix.

Rotations in Polar Coordinates

The rotation matrix $\mathbf{R}_\theta = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$
transforms any vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ to

$$\mathbf{R}_\theta \mathbf{x} = \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} x_1\cos\theta - x_2\sin\theta \\ x_1\sin\theta + x_2\cos\theta \end{pmatrix}$$

Introducing polar coordinates (r, η),
where $\mathbf{x} = (x_1, x_2) = r(\cos\eta, \sin\eta)$,
and then using trigonometric identities, we obtain

$$\mathbf{R}_\theta \mathbf{x} = r \begin{pmatrix} \cos\eta\cos\theta - \sin\eta\sin\theta \\ \cos\eta\sin\theta + \sin\eta\cos\theta \end{pmatrix} = r \begin{pmatrix} \cos(\eta+\theta) \\ \sin(\eta+\theta) \end{pmatrix}$$

This makes it easy to verify that:

1. $\mathbf{R}_\theta \mathbf{R}_\eta = \mathbf{R}_\eta \mathbf{R}_\theta = \mathbf{R}_{\theta+\eta}$ for all $\theta, \eta \in \mathbb{R}$;
2. $\mathbf{R}_{\theta+2k\pi} = \mathbf{R}_\theta$ for all $\theta \in \mathbb{R}$ and all $k \in \mathbb{Z}$.

Does a Rotation Matrix Have Real Eigenvalues?

The characteristic equation $|\mathbf{R}_\theta - \lambda \mathbf{I}| = 0$ takes the form

$$0 = \begin{vmatrix} \cos\theta - \lambda & -\sin\theta \\ \sin\theta & \cos\theta - \lambda \end{vmatrix} = (\cos\theta - \lambda)^2 + \sin^2\theta = 1 - 2\lambda\cos\theta + \lambda^2$$

1. A degenerate case occurs when $\theta = k\pi$ for some $k \in \mathbb{Z}$ so $\cos\theta = (-1)^k$ and $\sin\theta = 0$.

   Indeed $\mathbf{R}_\theta$ reduces to $(-1)^k \mathbf{I}_2$ in this degenerate case, where $\mathbf{I}_2$ denotes the identity matrix.
2. Otherwise, the real matrix $\mathbf{R}_\theta$ has no real eigenvalues.

   To show this, suppose that $\sin\theta \neq 0$ and so $\cos^2\theta < 1$.

   Then the characteristic equation $(\cos\theta - \lambda)^2 + \sin^2\theta = 0$ can be written as $(\cos\theta - \lambda)^2 = (\mathrm{i}\sin\theta)^2$.

   This has two distinct complex conjugate roots or eigenvalues given by $\lambda = \cos\theta \pm \mathrm{i}\sin\theta = e^{\pm\,\mathrm{i}\,\theta}$.

   The associated eigenspaces must both consist of complex eigenvectors.

Fundamental Theorem of Algebra

Theorem (Fundamental Theorem of Algebra)

Let $\mathbb{C} \ni \lambda \mapsto P(\lambda) = \lambda^n + \sum_{k=0}^{n-1} p_k \lambda^k$
be a polynomial function of λ *of degree* n*,*
possibly with complex coefficients p_k*.*
Then there exists at least one root $\hat{\lambda} \in \mathbb{C}$ *such that* $P(\hat{\lambda}) = 0$*.*

The proof, which is omitted from these notes,
involves relatively advanced techniques in complex analysis.

Polynomial Remainder Theorem

Taken from EMEA section 4.7, then extended to $\mathbb{C}$.

Theorem

Given the fraction $P(x)/Q(x)$,
let the numerator function $\mathbb{C} \ni x \mapsto P(x) \in \mathbb{C}$
and denominator function $\mathbb{C} \ni x \mapsto Q(x) \in \mathbb{C}$
be polynomials of degrees m and n respectively, where $m > n$.

Then there exist a quotient polynomial $q(x)$ of degree $m - n$,
as well as a remainder polynomial $r(x)$ of degree less than n,
such that $P(x) \equiv q(x)Q(x) + r(x)$.

Polynomial Remainder Theorem: Special Case

An important special case of the polynomial remainder theorem occurs when the denominator function takes the form $Q(x) \equiv x - c$ for some $c \in \mathbb{C}$, so that $n = 1$.

In this case the polynomial remainder theorem states that $P(x) \equiv q(x)(x - c) + r(x)$ where the remainder $r(x)$ is a polynomial of degree 0, so a constant $r \in \mathbb{C}$.

Thus $P(x) \equiv q(x)(x - c) + r$.

Putting $x = c$ gives $P(c) = r$.

Applying this result when $r = 0$ implies that the polynomial equation $P(x) = 0$ of degree m has a root $x = c$ if and only if $P(x) = q(x)(x - c)$, where $q(x)$ is of degree $m - 1$.

In other words, $x - c$ must be a factor of the polynomial $P(x)$.

Polynomial Factorization: Theorem

Theorem (Polynomial Factor Theorem)

Suppose that $\mathbb{C} \ni x \mapsto P(x) \in \mathbb{C}$
is a polynomial function of degree n.

Then $P(x) = \prod_{k=1}^{n}(x - c_k)$ *where the numbers* $c_1, \ldots, c_n \in \mathbb{C}$
are the n *roots, possibly coincident, of the equation* $P(x) = 0$.

Polynomial Factorization: Proof

Proof.

The proof will be by induction on n.

The result is trivial when $n = 1$.

For $n > 1$, by the fundamental theorem of algebra,
the equation $P(x) = 0$ has a root $c_n \in \mathbb{C}$.

The polynomial remainder theorem implies that there exists
a polynomial function $q(x)$ of degree $n - 1$
such that $P(x) = q(x)(x - c_n)$.

As the induction hypothesis, suppose that $q(x) = \prod_{k=1}^{n-1}(x - c_k)$.

It follows that $P(x) = q(x)(x - c_n) = \prod_{k=1}^{n}(x - c_k)$.

This confirms the induction step. □

Polynomial Factorization

Theorem

Given the $n \times n$ matrix $\mathbf{A}$,
the characteristic polynomial $P(\lambda) = |\mathbf{A} - \lambda\mathbf{I}$ of degree n
can be factorized
as the product $P_n(\lambda) \equiv \prod_{r=1}^{n}(\lambda - \lambda_r)$ of exactly n linear terms.

Characteristic Roots as Eigenvalues

Theorem

Every $n \times n$ matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ with complex elements has exactly n eigenvalues (real or complex) corresponding to the roots, counting multiple roots, of the characteristic equation $|\mathbf{A} - \lambda \mathbf{I}| = 0$.

Proof.

The characteristic equation can be written in the form $P_n(\lambda) = 0$ where $P_n(\lambda) \equiv |\lambda \mathbf{I} - \mathbf{A}|$ is the characteristic polynomial, which has degree n.

Because of polynomial factorization, the polynomial $|\lambda \mathbf{I} - \mathbf{A}|$ equals the product $\prod_{r=1}^{n} (\lambda - \lambda_r)$ of n linear terms, where each λ_r is a root of $P_n(\lambda) = 0$.

For any of these roots λ_r the matrix $\mathbf{A} - \lambda_r \mathbf{I}$ is singular.

So there exists $\mathbf{x} \neq \mathbf{0}$ such that $(\mathbf{A} - \lambda_r \mathbf{I})\mathbf{x} = \mathbf{0}$ or $\mathbf{A}\mathbf{x} = \lambda_r \mathbf{x}$, implying that λ_r is an eigenvalue. □

Outline

Linear Independence of Eigenvectors

The following theorem tells us
that eigenvectors associated with distinct eigenvalues
must be linearly independent.

Theorem

Let $\{\lambda_k\}_{k=1}^m = \{\lambda_1, \lambda_2, \ldots, \lambda_m\}$
be any collection of $m \leq n$ *distinct eigenvalues.*

Then any corresponding set $\{\mathbf{x}_k\}_{k=1}^m$ *of associated eigenvectors must be linearly independent.*

The proof will be by induction on m.

Because $\mathbf{x}_1 \neq \mathbf{0}$, the set $\{\mathbf{x}_1\}$ is linearly independent.

So the result is evidently true when $m = 1$.

As the induction hypothesis, suppose the result holds for $m - 1$.

Completing the Proof by Induction, I

Suppose that one solution of the equation $\mathbf{A}\mathbf{x} = \lambda_m \mathbf{x}$, which may be zero, is the linear combination $\mathbf{x} = \sum_{k=1}^{m-1} \alpha_k \mathbf{x}_k$ of the preceding $m-1$ eigenvectors. Hence

$$\mathbf{A}\mathbf{x} = \lambda_m \mathbf{x} = \sum_{k=1}^{m-1} \alpha_k \lambda_m \mathbf{x}_k$$

Then the hypothesis that $\{(\lambda_k, \mathbf{x}_k)\}_{k=1}^{m-1}$ is a collection of eigenpairs implies that this $\mathbf{x}$ satisfies

$$\mathbf{A}\mathbf{x} = \sum_{k=1}^{m-1} \alpha_k \mathbf{A}\mathbf{x}_k = \sum_{k=1}^{m-1} \alpha_k \lambda_k \mathbf{x}_k$$

Subtracting this equation from the prior equation gives

$$\mathbf{0} = \sum_{k=1}^{m-1} \alpha_k (\lambda_m - \lambda_k) \mathbf{x}_k$$

Completing the Proof by Induction, II

So we have

$$\mathbf{0} = \sum_{k=1}^{m-1} \alpha_k(\lambda_m - \lambda_k)\mathbf{x}_k$$

The induction hypothesis is that the set $\{\mathbf{x}_k\}_{k=1}^{m-1}$
of distinct eigenvectors is linearly independent, implying that

$$\alpha_k(\lambda_m - \lambda_k) = 0 \quad \text{for } k = 1, \ldots, m-1$$

But we are assuming
that all the m eigenvalues in $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ are distinct,
so $\lambda_m - \lambda_k \neq 0$ for $k = 1, \ldots, m-1$.

It follows that $\alpha_k = 0$ for $k = 1, \ldots, m-1$.

So we have proved that if $\mathbf{x} = \sum_{k=1}^{m-1} \alpha_k \mathbf{x}_k$ solves $\mathbf{A}\mathbf{x} = \lambda_m \mathbf{x}$,
then $\mathbf{x} = \mathbf{0}$, so $\mathbf{x}$ is not an eigenvector.

This completes the proof by induction that no eigenvector $\mathbf{x} \in E_{\lambda_m}$
can be a linear combination
of the eigenvectors $\mathbf{x}_k \in E_{\lambda_k}$ $(k = 1, \ldots, m-1)$. □

Outline

Similar Matrices

Definition

The two $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ are similar
just in case there exists an invertible $n \times n$ matrix $\mathbf{S}$
such that the following three equivalent statements all hold

$$\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S} \iff \mathbf{S}\mathbf{B} = \mathbf{A}\mathbf{S} \iff \mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$$

When $\mathbf{A}$ and $\mathbf{B}$ are similar, we write $\mathbf{A} \sim \mathbf{B}$.

Similarity versus Quadratic Form Invariance

Remark

Previously we proved that two symmetric $n \times n$ matrices $\mathbf{A}$ and $\mathbf{B}$ have the same definiteness, or satisfy quadratic form invariance, provided there exists an invertible $n \times n$ matrix $\mathbf{R}$ such that $\mathbf{B} = \mathbf{R}\mathbf{A}\mathbf{R}^\top$.

We have just defined $\mathbf{A}$ and $\mathbf{B}$ to be similar just in case there exists an $n \times n$ matrix $\mathbf{S}$ such that $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$.

The two definitions of similarity and quadratic form invariance are evidently different in general, even when $\mathbf{A}$ and $\mathbf{B}$ are both symmetric.

In case, however, there is an orthogonal matrix $\mathbf{P}$ such that $\mathbf{B} = \mathbf{P}\mathbf{A}\mathbf{P}^\top = \mathbf{P}\mathbf{A}\mathbf{P}^{-1}$, the two symmetric matrices $\mathbf{A}$ and $\mathbf{B}$ are both similar and quadratic form invariant.

Similarity is an Equivalence Relation

Theorem

The similarity relation is an equivalence relation — i.e., $\sim$ is:

reflexive $\mathbf{A} \sim \mathbf{A}$;

symmetric $\mathbf{A} \sim \mathbf{B} \iff \mathbf{B} \sim \mathbf{A}$;

transitive $\mathbf{A} \sim \mathbf{B}$ & $\mathbf{B} \sim \mathbf{C} \Longrightarrow \mathbf{A} \sim \mathbf{C}$

Proof.

The proofs that $\sim$ is reflexive and symmetric are elementary.

Suppose that $\mathbf{A} \sim \mathbf{B}$ and $\mathbf{B} \sim \mathbf{C}$.

By definition, there exist invertible matrices $\mathbf{S}$ and $\mathbf{T}$ such that $\mathbf{B} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ and $\mathbf{C} = \mathbf{T}^{-1}\mathbf{B}\mathbf{T}$.

Define $\mathbf{U} := \mathbf{S}\mathbf{T}$, which is invertible with $\mathbf{U}^{-1} = \mathbf{T}^{-1}\mathbf{S}^{-1}$.

Then $\mathbf{C} = \mathbf{T}^{-1}(\mathbf{S}^{-1}\mathbf{A}\mathbf{S})\mathbf{T} = (\mathbf{T}^{-1}\mathbf{S}^{-1})\mathbf{A}(\mathbf{S}\mathbf{T}) = \mathbf{U}^{-1}\mathbf{A}\mathbf{U}$.

So $\mathbf{A} \sim \mathbf{C}$. □

Similar Matrices Have Identical Spectra

Theorem

If $\mathbf{A} \sim \mathbf{B}$ *then* $\mathcal{S}_{\mathbf{A}} = \mathcal{S}_{\mathbf{B}}$.

Proof.

Suppose that $\mathbf{A} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}$ and that $(\lambda, \mathbf{x})$ is an eigenpair of $\mathbf{A}$.

Then $\mathbf{x} \neq \mathbf{0}$ solves $\mathbf{A}\mathbf{x} = \mathbf{S}\mathbf{B}\mathbf{S}^{-1}\mathbf{x} = \lambda\mathbf{x}$.

Premultiplying each side of the equation $\mathbf{S}\mathbf{B}\mathbf{S}^{-1}\mathbf{x} = \lambda\mathbf{x}$ by $\mathbf{S}^{-1}$, it follows that $\mathbf{y} := \mathbf{S}^{-1}\mathbf{x}$ solves $\mathbf{B}\mathbf{y} = \lambda\mathbf{y}$.

Moreover, because $\mathbf{S}^{-1}$ has the inverse $\mathbf{S}$, the equation $\mathbf{S}^{-1}\mathbf{x} = \mathbf{y}$ would have only the trivial solution $\mathbf{x} = \mathbf{S}\mathbf{y} = \mathbf{0}$ in case $\mathbf{y} = \mathbf{0}$.

Hence $\mathbf{y} \neq \mathbf{0}$, implying that $(\lambda, \mathbf{y})$ is an eigenpair of $\mathbf{B}$.

A symmetric argument shows
that if $(\lambda, \mathbf{y})$ is an eigenpair of $\mathbf{B} = \mathbf{S}^{-1}\mathbf{S}\mathbf{A}$,
then $(\lambda, \mathbf{S}\mathbf{y})$ is an eigenpair of $\mathbf{A}$. □

Diagonalization Theorem

Definition

An $n \times n$ matrix $\mathbf{A}$ is diagonalizable just in case
it is similar to a diagonal matrix $\mathbf{\Lambda} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

Then the $n \times n$ matrix $\mathbf{S}$ that satisfies $\mathbf{\Lambda} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$
is said to diagonalize $\mathbf{A}$.

Theorem

Given any diagonalizable $n \times n$ matrix $\mathbf{A}$:

1. *The columns of any matrix $\mathbf{S}$ that diagonalizes $\mathbf{A}$ must consist of n linearly independent eigenvectors of $\mathbf{A}$.*
2. *The matrix $\mathbf{A}$ is diagonalizable if and only if it has a set of n linearly independent eigenvectors.*
3. *The matrix $\mathbf{A}$ and its diagonalization $\mathbf{\Lambda} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ have the same set of eigenvalues.*

Proof of Diagonalization Theorem: Part 1

Suppose that $\mathbf{\Lambda} = \mathbf{S}^{-1}\mathbf{A}\mathbf{S}$ and so $\mathbf{A}\mathbf{S} = \mathbf{S}\mathbf{\Lambda}$,
where $\mathbf{A} = (a_{ij})^{n \times n}$, $\mathbf{S} = (s_{ij})^{n \times n}$, and $\mathbf{\Lambda} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

Then for each $i, k \in \mathbb{N}_n$, equating the elements
in row i and column k of the equal matrices $\mathbf{AS}$ and $\mathbf{S\Lambda}$
implies that $\sum_{j=1}^{n} a_{ij} s_{jk} = \sum_{j=1}^{n} s_{ij} \delta_{jk} \lambda_k = s_{ik} \lambda_k$.

It follows that $\mathbf{A}\mathbf{s}^k = \lambda_k \mathbf{s}^k$
where $\mathbf{s}^k = (s_{ik})_{i=1}^n$ denotes the kth column of the matrix $\mathbf{S}$.

Because the diagonalizing matrix $\mathbf{S}$ must be invertible:

- each column $\mathbf{s}^k$ must be non-zero, so an eigenvector of $\mathbf{A}$;
- the set $\{\mathbf{s}^k\}_{k=1}^n$ of all these n columns
  must be linearly independent. □

Proof of Diagonalization Theorem: Parts 2 and 3

Proof of Part 2: By part 1, if the diagonalizing matrix $\mathbf{S}$ exists, its columns must form a set of n linearly independent eigenvectors for the matrix $\mathbf{A}$.

Conversely, suppose that $\mathbf{A}$ does have a set $\{\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^n\}$ of n linearly independent eigenvectors, with $\mathbf{A}\mathbf{x}^k = \lambda_k \mathbf{x}^k$ for $k = 1, 2, \ldots, n$.

Now define $\mathbf{S}$ as the $n \times n$ matrix whose kth column is the eigenvector $\mathbf{x}^k$, for each $k = 1, 2, \ldots, n$.

Then it is easy to check that $\mathbf{AS} = \mathbf{S\Lambda}$ where $\mathbf{\Lambda} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. □

Proof of Part 3: By definition, the two matrices $\mathbf{A}$ and $\mathbf{\Lambda}$ are similar.

So they have the same spectrum of eigenvalues.

Outline

Complex Conjugate Matrices

Recall that any complex number $c \in \mathbb{C}$ can be expressed as $a + \mathrm{i}\, b$ with $a \in \mathbb{R}$ as the real part and $b \in \mathbb{R}$ as the imaginary part.

The complex conjugate of c is $\bar{c} = a - \mathrm{i}\, b$.

Note that $c\bar{c} = \bar{c}c = (a + \mathrm{i}\, b)(a - \mathrm{i}\, b) = a^2 + b^2 = |c|^2$,
where $|c|$ is the modulus of c.

Note that any $m \times n$ complex matrix $\mathbf{C} = (c_{ij})_{m \times n} \in \mathbb{C}^{m \times n}$
can be written as $\mathbf{A} + \mathrm{i}\, \mathbf{B}$,
where $\mathbf{A} = (a_{ij})_{m \times n}$ and $\mathbf{B} = (b_{ij})_{m \times n}$ are real $m \times n$ matrices
whose respective elements satisfy $c_{ij} = a_{ij} + \mathrm{i}\, b_{ij}$ for all $i, j \in \mathbb{N}_n$.

The complex conjugate of the complex matrix $\mathbf{C} = \mathbf{A} + \mathrm{i}\, \mathbf{B}$
is $\bar{\mathbf{C}} = \mathbf{A} - \mathrm{i}\, \mathbf{B}$ with respective elements $\bar{c}_{ij} = a_{ij} - \mathrm{i}\, b_{ij}$
for all $i, j \in \mathbb{N}_n$.

Adjoint Matrices

The adjoint of the $m \times n$ complex matrix $\mathbf{C} = \mathbf{A} + \mathrm{i}\,\mathbf{B}$
is the $n \times m$ complex matrix $\mathbf{C}^* := \bar{\mathbf{C}}^\top = (\mathbf{A} - \mathrm{i}\,\mathbf{B})^\top = \mathbf{A}^\top - \mathrm{i}\,\mathbf{B}^\top$.

This is the transpose of the complex conjugate matrix $\bar{\mathbf{C}} = \mathbf{A} - \mathrm{i}\,\mathbf{B}$
whose elements are the complex conjugates $\bar{c}_{ij} = a_{ij} - \mathrm{i}\,b_{ij}$
of the corresponding elements $c_{ij} = a_{ij} + \mathrm{i}\,b_{ij}$ of $\mathbf{C}$.

That is, each element of $\mathbf{C}^*$ is given by $c^*_{ij} = a_{ji} - \mathrm{i}\,b_{ji}$.

Alternatively, the adjoint matrix $\mathbf{C}^* = \mathbf{A}^\top - \mathrm{i}\,\mathbf{B}^\top$
is the complex conjugate of the transpose matrix $\mathbf{C}^\top = \mathbf{A}^\top + \mathrm{i}\,\mathbf{B}^\top$.

In the case of a real matrix $\mathbf{A}$, whose imaginary part is $\mathbf{0}$,
its adjoint is simply the transpose $\mathbf{A}^\top$.

Self-Adjoint and Symmetric Matrices

An $n \times n$ complex matrix $\mathbf{C} = \mathbf{A} + i\mathbf{B}$
is defined to be self-adjoint just in case $\mathbf{C}^* = \mathbf{C}$
— that is, just in case $\mathbf{C}$ equals its own adjoint $\mathbf{C}^*$.

Self-adjointness holds if and only if $\mathbf{A}^\top - i\mathbf{B}^\top = \mathbf{A} + i\mathbf{B}$,
and so if and only if:

- the real part $\mathbf{A}$ is symmetric;
- the imaginary part $\mathbf{B}$ is anti-symmetric
  in the sense that $\mathbf{B}^\top = -\mathbf{B}$.

Of course, a real matrix is self-adjoint if and only if it is symmetric.

Outline

A Self-Adjoint Matrix has only Real Eigenvalues

Theorem

Any eigenvalue of a self-adjoint complex matrix is a real scalar.

Proof.

Given any $\mathbf{A} \in \mathbb{C}^{n \times n}$,
suppose that the scalar $\lambda \in \mathbb{C}$ and vector $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$
together satisfy the eigenvalue equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$.

Taking the adjoint of this equation gives $\mathbf{x}^*\mathbf{A}^* = \bar{\lambda}\mathbf{x}^*$.

By the associative law of complex matrix multiplication,
one has $\mathbf{x}^*\mathbf{A}\mathbf{x} = \mathbf{x}^*(\mathbf{A}\mathbf{x}) = \mathbf{x}^*(\lambda\mathbf{x}) = \lambda(\mathbf{x}^*\mathbf{x})$
as well as $\mathbf{x}^*\mathbf{A}^*\mathbf{x} = (\mathbf{x}^*\mathbf{A}^*)\mathbf{x} = (\bar{\lambda}\mathbf{x}^*)\mathbf{x} = \bar{\lambda}(\mathbf{x}^*\mathbf{x})$.

In case $\mathbf{A}$ is self-adjoint and so $\mathbf{A}^* = \mathbf{A}$,
it follows that $\lambda(\mathbf{x}^*\mathbf{x}) = \mathbf{x}^*\mathbf{A}\mathbf{x} = \mathbf{x}^*\mathbf{A}^*\mathbf{x} = \bar{\lambda}(\mathbf{x}^*\mathbf{x})$.

But $\mathbf{x}$ is an eigenvector, so $\mathbf{x} \neq \mathbf{0}$ and $\mathbf{x}^*\mathbf{x} = \sum_{j=1}^n \bar{x}_j x_j > 0$.

It follows that $\lambda = \bar{\lambda}$, so λ must be real. □

The Case of a Symmetric Real Matrix

Corollary

Given any symmetric real $n \times n$ matrix $\mathbf{A}$:

1. *any eigenvalue is a real scalar;*
2. *associated with any eigenvalue,*
   there must be at least one real eigenvector.

Proof.

1. The matrix $\mathbf{A}$ is trivially self-adjoint, so any eigenvalue is real.
2. Suppose that $\mathbf{a}, \mathbf{b} \in \mathbb{R}^n$ and $\lambda \in \mathbb{R}$
   satisfy $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$ where $\mathbf{x} = \mathbf{a} + \mathrm{i}\,\mathbf{b} \neq \mathbf{0} + \mathrm{i}\,\mathbf{0}$.

   Equating the real and imaginary parts of $\mathbf{A}\mathbf{x}$ and $\lambda\mathbf{x}$,
   it follows that $\mathbf{A}\mathbf{a} = \lambda\mathbf{a}$ and $\mathbf{A}\mathbf{b} = \lambda\mathbf{b}$.

   Also, at least one of $\mathbf{a}$ and $\mathbf{b}$ must be non-zero.

   So at least one of $\mathbf{a}$ and $\mathbf{b}$ is a real eigenvector
   that is associated with the eigenvalue λ. □

Outline

The Rayleigh Quotient Function

For all $\mathbf{x} \neq \mathbf{0}$, define the Rayleigh quotient function

$$\mathbb{R}^n \setminus \{\mathbf{0}\} \ni \mathbf{x} \mapsto f(\mathbf{x}) := \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \frac{\sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j}{\sum_{i=1}^n x_i^2}$$

It is homogeneous of degree zero, and left undefined at $\mathbf{x} = \mathbf{0}$.

Its partial derivative w.r.t. any component x_h of the vector $\mathbf{x}$ is

$$\frac{\partial f}{\partial x_h} = \frac{2}{(\mathbf{x}^\top \mathbf{x})^2} \left[\sum_{j=1}^n a_{hj} x_j (\mathbf{x}^\top \mathbf{x}) - (\mathbf{x}^\top \mathbf{A} \mathbf{x}) x_h \right]$$

The point $\hat{\mathbf{x}} \neq \mathbf{0}$ is critical if and only if $\partial f / \partial x_h = 0$ for all h,
and so if and only if $(\hat{\mathbf{x}}^\top \hat{\mathbf{x}}) \mathbf{A} \hat{\mathbf{x}} = (\hat{\mathbf{x}}^\top \mathbf{A} \hat{\mathbf{x}}) \hat{\mathbf{x}}$,
or equivalently, iff $\mathbf{A} \hat{\mathbf{x}} = \lambda \hat{\mathbf{x}}$ where $\lambda = (\hat{\mathbf{x}}^\top \mathbf{A} \hat{\mathbf{x}}) / (\hat{\mathbf{x}}^\top \hat{\mathbf{x}}) = f(\hat{\mathbf{x}})$.

That is, a point $\hat{\mathbf{x}} \neq \mathbf{0}$ is critical if and only if it is an eigenvector,
with the corresponding function value $f(\hat{\mathbf{x}})$
as the associated eigenvalue.

More Properties of the Rayleigh Quotient Function

Using the Rayleigh quotient function

$$\mathbb{R}^n \setminus \{\mathbf{0}\} \ni \mathbf{x} \mapsto f(\mathbf{x}) := \frac{\mathbf{x}^\top \mathbf{A} \mathbf{x}}{\mathbf{x}^\top \mathbf{x}} = \frac{\sum_{i=1}^n \sum_{j=1}^n x_i a_{ij} x_j}{\sum_{i=1}^n x_i^2}$$

one can state and prove the following lemma.

Lemma

Every $n \times n$ symmetric square matrix $\mathbf{A}$:

1. *has a maximum eigenvalue λ^* with λ^* real, and any associated eigenvector $\mathbf{x}^*$ as a maximum point of f;*
2. *has a minimum eigenvalue λ_* with λ_* real, and any associated eigenvector $\mathbf{x}_*$ as a minimum point of f;*
3. *satisfies $\mathbf{A} = \lambda \mathbf{I}$ if and only if $\lambda^* = \lambda_* = \lambda$.*

Proof of Parts 1 and 2

The unit sphere S^{n-1} is a closed and bounded subset of $\mathbb{R}^n$.

Moreover, the Rayleigh quotient function f is continuous when restricted to S^{n-1}.

By the extreme value theorem, f restricted to S^{n-1} must have:

- a maximum value λ^* attained at some point $\mathbf{x}^*$;
- a minimum value λ_* attained at some point $\mathbf{x}_*$.

Because f is homogeneous of degree zero,
these are the maximum and minimum values of f
over the whole domain $\mathbb{R}^n \setminus \{\mathbf{0}\}$.

In particular, f must be critical at any maximum point $\mathbf{x}^*$,
as well as at any minimum point $\mathbf{x}_*$.

But critical points must be eigenvectors.

This proves parts 1 and 2 of the lemma.

Part 3 is left as an exercise. □

The Eigenvalue Test of Definiteness

Theorem

The quadratic form $\mathbf{x}^\top \mathbf{A} \mathbf{x}$ *is:*

positive definite if and only if all eigenvalues of $\mathbf{A}$ *are positive;*

negative definite if and only if all eigenvalues of $\mathbf{A}$ *are negative;*

positive semi-definite if and only if all eigenvalues of $\mathbf{A}$ *are non-negative;*

negative semi-definite if and only if all eigenvalues of $\mathbf{A}$ *are non-positive;*

indefinite if and only if $\mathbf{A}$ *has both positive and negative eigenvalues.*

Proof.

The maximum and minimum values of the Rayleigh quotient function are respectively equal to the largest and smallest eigenvalues λ^* and λ_*.

The rest of the proof is straightforward. □

Final Remark on Comparing Definiteness Tests

Remark

The eigenvalue test features prominently in Math Econ textbooks, including EMEA and FMEA.

Yet there are no simple finite algorithms for finding eigenvalues, or critical points of the Rayleigh quotient function.

For this reason, the symmetric pivoting algorithm discussed earlier is likely to be much more practical if there are more than about 3 dimensions.

Symmetric pivoting, of course, is also computationally superior to the Sylvester criterion which requires calculating many determinants.

The diagonalization procedure we are about to discuss could also be used to find the eigenvalues.

But it relies on finding successive eigenvalues by some other method, yet to be specified.

Outline

Orthogonal and Orthonormal Sets of Vectors

Recall our earlier definition:

Definition

A set of k vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \subset \mathbb{R}^n$ is said to be:

- pairwise orthogonal just in case $\mathbf{x}_i \cdot \mathbf{x}_j = 0$ whenever $j \neq i$;
- orthonormal just in case, in addition, each $\|\mathbf{x}_i\| = 1$
  — i.e., all k elements of the set are vectors of unit length.

The set of k vectors $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\} \subset \mathbb{R}^n$ is orthonormal
just in case $\mathbf{x}_i \cdot \mathbf{x}_j = \delta_{ij}$ for all pairs $i, j \in \{1, 2, \ldots, k\}$.

Theorem Characterizing Orthogonal Matrices: Recall

Definition

Any $n \times n$ matrix is **orthogonal** just in case its n columns (or rows) form an orthonormal set.

Theorem

Given any $n \times n$ matrix $\mathbf{P}$, the following are equivalent:

1. $\mathbf{P}$ *is orthogonal;*
2. $\mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top\mathbf{P} = \mathbf{I}$*;*
3. $\mathbf{P}^{-1} = \mathbf{P}^\top$*;*
4. $\mathbf{P}^\top$ *is orthogonal.*

Proofs of Fourfold Equivalence

$(\mathbf{1}) \iff (\mathbf{2})$: Each (i,j) element of the two matrices satisfies

$$(\mathbf{P}\mathbf{P}^\top)_{ij} = \sum\nolimits_{k=1}^n p_{ik} p_{jk} = \mathbf{p}_i^\top \cdot \mathbf{p}_j^\top$$
$$\text{and} \quad (\mathbf{P}^\top\mathbf{P})_{ij} = \sum\nolimits_{k=1}^n p_{ki} p_{kj} = \mathbf{p}_i \cdot \mathbf{p}_j$$

Both right-hand sides equal the Kronecker delta δ_{ij}
if and only if the rows and columns both form orthonormal sets,
as well as if and only if $\mathbf{P}\mathbf{P}^\top = \mathbf{P}^\top\mathbf{P} = \mathbf{I}$.

$(\mathbf{2}) \iff (\mathbf{3})$: Immediate from the definition of inverse matrix.

$(\mathbf{3}) \iff (\mathbf{4})$: Taking the transpose of (3), then premultiplying each side by the invertible matrix $(\mathbf{P}^{-1})^\top = (\mathbf{P}^\top)^{-1}$, one has

$$\mathbf{P}^{-1} = \mathbf{P}^\top \iff (\mathbf{P}^{-1})^\top \mathbf{P}^{-1} = (\mathbf{P}^\top)^{-1}\mathbf{P}^\top = \mathbf{I}$$

This proves that $\mathbf{P}^{-1} = \mathbf{P}^\top$ if and only if $\mathbf{P}^\top$ is orthogonal. □

The Complex Case: Self-Adjoint and Unitary Matrices

We briefly consider matrices with complex elements.

Recall that the adjoint $\mathbf{A}^*$ of an $m \times n$ matrix $\mathbf{A}$
is the matrix formed from the transpose $\mathbf{A}^\top$
by taking the complex conjugate of each element.

The appropriate extension to complex numbers of:

- a symmetric matrix satisfying $\mathbf{A}^\top = \mathbf{A}$
  is a self-adjoint matrix satisfying $\mathbf{A}^* = \mathbf{A}$;
- an orthogonal matrix satisfying $\mathbf{P}^{-1} = \mathbf{P}^\top$
  is a unitary matrix satisfying $\mathbf{U}^{-1} = \mathbf{U}^*$.

Orthogonal Projections

Definition

A general $n \times n$ matrix $\mathbf{P}$ is idempotent just in case $\mathbf{P}^2 = \mathbf{P}$, and so $\mathbf{P}^n = \mathbf{P}$ for all $n \in \mathbb{N}$.

A symmetric $n \times n$ matrix $\mathbf{P}$ is an orthogonal projection just in case it is idempotent.

It follows that if $\mathbf{P}$ is an orthogonal projection, then $\mathbf{P} - \mathbf{P}^2 = \mathbf{P}(\mathbf{I} - \mathbf{P}) = \mathbf{0}_{n \times n}$.

So the projection $\mathbf{y} = \mathbf{P}\mathbf{x}$ and the displacement $\mathbf{z} = \mathbf{x} - \mathbf{P}\mathbf{x}$ of any $x \in \mathbb{R}^n$ are orthogonal because

$$\mathbf{y}^\top \mathbf{z} = (\mathbf{P}\mathbf{x})^\top (\mathbf{x} - \mathbf{P}\mathbf{x}) = \mathbf{x}^\top \mathbf{P}\mathbf{x} - \mathbf{x}^\top \mathbf{P}^2 \mathbf{x} = \mathbf{x}^\top (\mathbf{P} - \mathbf{P}^2)\mathbf{x} = 0$$

Remark

An orthogonal projection matrix is generally not orthogonal.

Orthogonal Complements

Example

Given any $m, n \in \mathbb{N}$ with $m < n$,
the $m \times n$ block diagonal matrix $\mathbf{diag}(\mathbf{I}_m, \mathbf{0}_{m\times(n-m)})$
is a projection from $\mathbb{R}^n$ onto $\mathbb{R}^m$.

Definition

A subset $L \subseteq \mathbb{R}^n$ is a linear subspace just in case $\lambda\mathbf{x} + \mu\mathbf{y} \in L$
for every pair of vectors $\mathbf{x}, \mathbf{y}$ in L and every pair of scalars λ, μ in $\mathbb{R}$.

Definition

Given any linear subspace L of $\mathbb{R}^n$, its orthogonal complement $L^\perp$
is the set of all vectors $\mathbf{y} \in \mathbb{R}^n$ such that $\mathbf{x} \cdot \mathbf{y} = 0$ for all $\mathbf{x} \in L$.

Two Examples

Example

Suppose that L is the space spanned
by any finite subset $\{\mathbf{e}^i \mid i \in I\}$
of the canonical basis $\{\mathbf{e}^i \mid i = 1, 2, \ldots, n\}$ of $\mathbb{R}^n$.

Then $L^\perp$ is the space spanned
by the complementary set $\{\mathbf{e}^i \mid i \notin I\}$ of canonical basis vectors.

Example

Any $\mathbf{c} \neq \mathbf{0}$ in $\mathbb{R}^n$ generates the straight line $L(\mathbf{c}) = \{\lambda \mathbf{c} \mid \lambda \in \mathbb{R}\}$,
which is a one-dimensional linear subspace in $\mathbb{R}^n$.

Its orthogonal complement $L(\mathbf{c})^\perp = \{\mathbf{x} \mid \mathbf{c} \cdot \mathbf{x} = 0\}$
consists of an $n - 1$-dimensional subspace
which is the unique hyperplane in $\mathbb{R}^n$ that has $\mathbf{c}$ as a normal.

Useful Lemma

Lemma

Suppose that the $n \times m$ matrix $\mathbf{X}$ has full rank $m < n$.
Then $\mathbf{X}^\top \mathbf{X}$ is invertible.

Proof.

Because $\mathbf{X}$ has full rank, its m columns are linearly independent.
It follows that $\mathbf{X}\mathbf{y} = \mathbf{0} \implies \mathbf{y} = \mathbf{0}$.
But then

$$\mathbf{X}^\top \mathbf{X}\mathbf{y} = \mathbf{0} \implies \mathbf{y}^\top \mathbf{X}^\top \mathbf{X}\mathbf{y} = (\mathbf{X}\mathbf{y})^\top \mathbf{X}\mathbf{y} = \mathbf{0} \implies \mathbf{X}\mathbf{y} = \mathbf{0}$$

So

$$\mathbf{X}^\top \mathbf{X}\mathbf{y} = \mathbf{0} \implies \mathbf{y} = \mathbf{0}$$

□

Orthogonal Projection Matrices

Theorem

Suppose that the $n \times m$ matrix $\mathbf{X}$ has full rank $m < n$.

Let $L \subset \mathbb{R}^n$ be the linear subspace spanned by m linearly independent columns of $\mathbf{X}$.

Define the $n \times n$ matrix $\mathbf{P} := \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$. Then:

1. *The matrix $\mathbf{P}$ is an orthogonal projection onto L.*
2. *The matrix $\mathbf{I} - \mathbf{P}$ is an orthogonal projection onto the orthogonal complement $L^\perp$ of L.*
3. *For each vector $\mathbf{y} \in \mathbb{R}^n$, its orthogonal projection onto L is the unique vector $\mathbf{v} = \mathbf{P}\mathbf{y} \in L$ that minimizes the distance $\|\mathbf{y} - \mathbf{v}\|$ between $\mathbf{y}$ and L — i.e., $\|\mathbf{y} - \mathbf{v}\| \leq \|\mathbf{y} - \mathbf{u}\|$ for all $\mathbf{u} \in L$.*

Proof of Part 1

First note that if $\mathbf{X}$ is an $n \times m$ matrix, then $\mathbf{X}^\top \mathbf{X}$ is $m \times m$.

Then, provided that $(\mathbf{X}^\top \mathbf{X})^{-1}$ exists,
so does the $n \times n$ matrix $\mathbf{P} = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$.

Because of the rules for the transposes of products and inverses, the definition $\mathbf{P} := \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top$ implies that $\mathbf{P}^\top = \mathbf{P}$ and also

$$\mathbf{P}^2 = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{X}(\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top = \mathbf{P}$$

Moreover, if $\mathbf{P}\mathbf{v} = \mathbf{0}$ and $\mathbf{u} = \mathbf{P}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \mathbf{x}^\top \mathbf{P}^\top \mathbf{v} = \mathbf{x}^\top \mathbf{P}\mathbf{v} = 0$$

Finally, for every $\mathbf{y} \in \mathbb{R}^n$, the vector $\mathbf{P}\mathbf{y}$ equals $\mathbf{X}\mathbf{b}$, where

$$\mathbf{b} = (\mathbf{X}^\top \mathbf{X})^{-1}\mathbf{X}^\top \mathbf{y}$$

Hence $\mathbf{P}\mathbf{y} \in L$. □

Proof of Part 2

Evidently $(\mathbf{I} - \mathbf{P})^\top = \mathbf{I} - \mathbf{P}^\top = \mathbf{I} - \mathbf{P}$, and

$$(\mathbf{I} - \mathbf{P})^2 = \mathbf{I} - 2\mathbf{P} + \mathbf{P}^2 = \mathbf{I} - 2\mathbf{P} + \mathbf{P} = \mathbf{I} - \mathbf{P}$$

Hence $\mathbf{I} - \mathbf{P}$ is a projection.

This projection is also orthogonal because if $(\mathbf{I} - \mathbf{P})\mathbf{v} = \mathbf{0}$ and $\mathbf{u} = (\mathbf{I} - \mathbf{P})\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$, then

$$\mathbf{u} \cdot \mathbf{v} = \mathbf{u}^\top \mathbf{v} = \mathbf{x}^\top (\mathbf{I} - \mathbf{P})^\top \mathbf{v} = \mathbf{x}^\top (\mathbf{I} - \mathbf{P})\mathbf{v} = 0$$

Next, suppose that $\mathbf{v} = \mathbf{X}\mathbf{b} \in L$ and that $\mathbf{y} = (\mathbf{I} - \mathbf{P})\mathbf{x}$ belongs to the range of $(\mathbf{I} - \mathbf{P})$. Then

$$\mathbf{y} \cdot \mathbf{v} = \mathbf{y}^\top \mathbf{v} = \mathbf{x}^\top (\mathbf{I} - \mathbf{P})^\top \mathbf{X}\mathbf{b} = \mathbf{x}^\top \mathbf{X}\mathbf{b} - \mathbf{x}^\top \mathbf{X}\mathbf{b} = 0$$

Hence $\mathbf{y} \in L^\perp$. □

Proof of Part 3

For any vector $\mathbf{v} = \mathbf{X}\mathbf{b} \in L$ and any $\mathbf{y} \in \mathbb{R}^n$,
because $\mathbf{y}^\top \mathbf{X}\mathbf{b}$ and $\mathbf{b}^\top \mathbf{X}^\top \mathbf{y}$ are equal scalars, one has

$$\|\mathbf{y} - \mathbf{v}\|^2 = (\mathbf{y} - \mathbf{X}\mathbf{b})^\top (\mathbf{y} - \mathbf{X}\mathbf{b}) = \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\mathbf{b} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{X}\mathbf{b}$$

Now define $\hat{\mathbf{b}} := (\mathbf{X}^\top \mathbf{X})^{-1} \mathbf{X}^\top \mathbf{y}$ (which is the OLS estimator of $\mathbf{b}$ in the linear regression equation $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$) and $\hat{\mathbf{v}} := \mathbf{X}\hat{\mathbf{b}} = \mathbf{P}\mathbf{y}$.

Because $\mathbf{P}^\top \mathbf{P} = \mathbf{P}^\top = \mathbf{P} = \mathbf{P}^2$, one has

$$\begin{aligned}
\|\mathbf{y} - \mathbf{v}\|^2 &= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{X}\mathbf{b} + \mathbf{b}^\top \mathbf{X}^\top \mathbf{X}\mathbf{b} \\
&= (\mathbf{b} - \hat{\mathbf{b}})^\top \mathbf{X}^\top \mathbf{X}(\mathbf{b} - \hat{\mathbf{b}}) + \mathbf{y}^\top \mathbf{y} - \hat{\mathbf{b}}^\top \mathbf{X}^\top \mathbf{X}\hat{\mathbf{b}} \\
&= \|\mathbf{v} - \hat{\mathbf{v}}\|^2 + \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}^\top \mathbf{P}\mathbf{y} = \|\mathbf{v} - \hat{\mathbf{v}}\|^2 + \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y}
\end{aligned}$$

On the other hand, given that $\hat{\mathbf{v}} = \mathbf{P}\mathbf{y}$, one also has

$$\begin{aligned}
\|\mathbf{y} - \hat{\mathbf{v}}\|^2 &= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \hat{\mathbf{v}} + \hat{\mathbf{v}}^\top \hat{\mathbf{v}} \\
&= \mathbf{y}^\top \mathbf{y} - 2\mathbf{y}^\top \mathbf{P}\mathbf{y} + \mathbf{y}^\top \mathbf{P}^\top \mathbf{P}\mathbf{y} = \mathbf{y}^\top \mathbf{y} - \mathbf{y}^\top \mathbf{P}\mathbf{y}
\end{aligned}$$

So $\|\mathbf{y} - \mathbf{v}\|^2 - \|\mathbf{y} - \hat{\mathbf{v}}\|^2 = \|\mathbf{v} - \hat{\mathbf{v}}\|^2 \geq 0$ with $=$ iff $\mathbf{v} = \hat{\mathbf{v}}$. □

Outline

A Minor Lemma

Lemma

Let $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix.*

Suppose that λ *and* μ *are distinct eigenvalues, with corresponding eigenvectors* $\mathbf{x}$ *and* $\mathbf{y}$*.*

Then $\mathbf{x}$ *and* $\mathbf{y}$ *are orthogonal — that is,* $\mathbf{x} \cdot \mathbf{y} = 0$*.*

Proof.

Suppose that the non-zero vectors $\mathbf{x}$ and $\mathbf{y}$ satisfy $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ and $\mathbf{A}\mathbf{y} = \mu\mathbf{y}$.

Because $\mathbf{A}$ is symmetric, one has

$$\lambda\mathbf{x}^\top\mathbf{y} = (\mathbf{A}\mathbf{x})^\top\mathbf{y} = \mathbf{x}^\top\mathbf{A}^\top\mathbf{y} = \mathbf{x}^\top\mathbf{A}\mathbf{y} = \mu\mathbf{x}^\top\mathbf{y}$$

In case $\lambda \neq \mu$, it follows that $0 = \mathbf{x}^\top\mathbf{y} = \mathbf{x} \cdot \mathbf{y}$. □

A Useful Lemma

Lemma

Let $\mathbf{A}$ *be a symmetric* $n \times n$ *matrix.*

Suppose that there are $m < n$ *eigenvectors* $\{\mathbf{u}_k\}_{k=1}^m$ *which form an orthonormal set of column* n*-vectors, as well as the* m *columns of an* $n \times m$ *matrix* $\mathbf{U}$*.*

Then there is at least one more eigenvector $\mathbf{x} \neq \mathbf{0}_n$ *that satisfies* $\mathbf{U}^\top \mathbf{x} = \mathbf{0}_n$
— i.e., it is orthogonal to each of the m *eigenvectors* $\mathbf{u}_k$*.*

Constructive Proof, Part 1

For each eigenvector $\mathbf{u}_k$, let λ_k be the associated eigenvalue, so that $\mathbf{A}\mathbf{u}_k = \lambda_k \mathbf{u}_k$ for $k = 1, 2, \ldots, m$.

Then the $n \times m$ matrix $\mathbf{U}$ satisfies $\mathbf{AU} = \mathbf{U\Lambda}$ where $\mathbf{\Lambda}$ is the $m \times m$ matrix $\mathbf{diag}(\lambda_k)_{k=1}^m$.

Also, because the eigenvectors $\{\mathbf{u}_k\}_{k=1}^m$ form an orthonormal set, one has $\mathbf{U}^\top \mathbf{U} = \mathbf{I}_m$.

Hence $\mathbf{U}^\top \mathbf{AU} = \mathbf{U}^\top \mathbf{U\Lambda} = \mathbf{\Lambda}$.

Also, transposing $\mathbf{AU} = \mathbf{U\Lambda}$ gives $\mathbf{U}^\top \mathbf{A} = \mathbf{\Lambda U}^\top$.

Constructive Proof, Part 2

Consider now the $n \times n$ matrix $\hat{\mathbf{A}} := (\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)\mathbf{A}(\mathbf{I}_n - \mathbf{U}\mathbf{U}^\top)$.

Then $\hat{\mathbf{A}}$ is symmetric because both $\mathbf{A}$ and $\mathbf{U}\mathbf{U}^\top$ are symmetric.

Note that, because $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$ and so $\mathbf{U}^\top\mathbf{A} = \mathbf{\Lambda}\mathbf{U}^\top$, one has

$$\begin{aligned}\hat{\mathbf{A}} &= \mathbf{A} - \mathbf{U}\mathbf{U}^\top\mathbf{A} - \mathbf{A}\mathbf{U}\mathbf{U}^\top + \mathbf{U}\mathbf{U}^\top\mathbf{A}\mathbf{U}\mathbf{U}^\top \\ &= \mathbf{A} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top + \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top = \mathbf{A} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\end{aligned}$$

Because the matrix $\hat{\mathbf{A}}$ is symmetric,
it at least one real eigenvalue λ.

Let $\mathbf{x} \neq \mathbf{0}$ be an associated real eigenvector, which must satisfy

$$\hat{\mathbf{A}}\mathbf{x} = (\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\mathbf{A}(\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\mathbf{x} = (\mathbf{A} - \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top)\mathbf{x} = \lambda\mathbf{x}$$

Pre-multiplying each side of the last equality
by the $m \times n$ matrix $\mathbf{U}^\top$ shows that

$$\lambda\mathbf{U}^\top\mathbf{x} = \mathbf{U}^\top\mathbf{A}\mathbf{x} - \mathbf{U}^\top\mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top\mathbf{x} = \mathbf{\Lambda}\mathbf{U}^\top\mathbf{x} - \mathbf{\Lambda}\mathbf{U}^\top\mathbf{x} = \mathbf{0}_m$$

Constructive Proof, Part 3

There are now two cases.

Consider first the generic case
when $\hat{\mathbf{A}}$ has at least one eigenvalue $\lambda \neq 0$.

Then there is a corresponding eigenvector $\mathbf{x} \neq \mathbf{0}$ of $\hat{\mathbf{A}}$
that was shown in Part 2 to satisfy $\lambda \mathbf{U}^\top \mathbf{x} = \mathbf{0}_m$ and so $\mathbf{U}^\top \mathbf{x} = \mathbf{0}_m^\top$.

But then the earlier equation

$$\hat{\mathbf{A}}\mathbf{x} = (\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\mathbf{A}(\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\mathbf{x} = (\mathbf{A} - \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top)\mathbf{x} = \lambda \mathbf{x}$$

implies that

$$\mathbf{A}\mathbf{x} = (\hat{\mathbf{A}} + \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top)\mathbf{x} = \hat{\mathbf{A}}\mathbf{x} = \lambda \mathbf{x}$$

Hence $\mathbf{x}$ is an eigenvector of $\mathbf{A}$ as well as of $\hat{\mathbf{A}}$.

Constructive Proof, Part 4

The remaining exceptional case occurs
when the only eigenvalue of the symmetric matrix $\hat{\mathbf{A}}$ is $\lambda = 0$.

Given the properties of the Rayleigh quotient function,
this implies that $\hat{\mathbf{A}} = \mathbf{0}$ and so $\mathbf{A} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top$.

Then any vector $\mathbf{x} \neq \mathbf{0}$ satisfying $\mathbf{U}^\top\mathbf{x} = \mathbf{0}$
must satisfy $\mathbf{A}\mathbf{x} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^\top\mathbf{x} = \mathbf{0}$.

This implies that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$
associated with the eigenvalue $\lambda = 0$.

So, whether or not the only eigenvalue of $\hat{\mathbf{A}}$ is 0,
associated with the common eigenvalue λ of both $\hat{\mathbf{A}}$ and $\mathbf{A}$
is an eigenvector $\mathbf{x}$ of $\mathbf{A}$ that satisfies $\mathbf{U}^\top\mathbf{x} = \mathbf{0}_m^\top$.

This completes the proof. □

Spectral Theorem

Theorem

Given any symmetric $n \times n$ matrix $\mathbf{A}$:

1. *the set of all its eigenvectors spans the whole of $\mathbb{R}^n$;*
2. *there exists an orthogonal matrix $\mathbf{P}$ that diagonalizes $\mathbf{A}$ in the sense that $\mathbf{P}^\top \mathbf{A} \mathbf{P} = \mathbf{P}^{-1} \mathbf{A} \mathbf{P}$ is a diagonal matrix $\mathbf{\Lambda}$, whose elements are the eigenvalues of $\mathbf{A}$, all real.*

When $\mathbf{A}$ is a complex $n \times n$ matrix
that is required to be self-adjoint rather than symmetric,
and the eigenvectors may be complex, the corresponding result is:

1. the set of all eigenvectors of $\mathbf{A}$ spans the whole of $\mathbb{C}^n$;
2. there exists a unitary matrix $\mathbf{U}$ that diagonalizes $\mathbf{A}$ in the sense that $\mathbf{U}^* \mathbf{A} \mathbf{U} = \mathbf{U}^{-1} \mathbf{A} \mathbf{U}$ is a diagonal matrix $\mathbf{\Lambda}$ whose elements are the eigenvalues of $\mathbf{A}$, all real.

We give a proof for the case when $\mathbf{A}$ is real and symmetric.

Proof of Spectral Theorem, Part 1

The symmetric matrix $\mathbf{A}$ has at least one eigenvalue λ, which must be real.

The associated eigenvector $\mathbf{x}$, normalized to satisfy $\mathbf{x}^\top \mathbf{x} = 1$, forms an orthonormal set $\{\mathbf{u}_1\}$ consisting of only one n-vector.

As the induction hypothesis,
suppose that there are $m < n$ eigenvectors $\{\mathbf{u}_k\}_{k=1}^m$
which form an orthonormal set of m vectors.

We have just proved that this hypothesis holds for $m = 1$.

The "useful lemma" establishes the induction step showing that, if the hypothesis holds for any $m \in \mathbb{N}_{n-1}$, then it holds for $m + 1$.

So the result follows for $m = n$ by induction.

In particular, when $m = n$, there exists an orthonormal set of n eigenvectors, which must then span the whole of $\mathbb{R}^n$.

Proof of Spectral Theorem, Part 2

Also, by the previous result,
we can take $\mathbf{P}$ as an orthogonal matrix
whose columns are an orthonormal set of n eigenvectors.

Then $\mathbf{AP} = \mathbf{P}\boldsymbol{\Lambda}$.

So premultiplying by $\mathbf{P}^\top = \mathbf{P}^{-1}$ gives $\mathbf{P}^\top\mathbf{AP} = \mathbf{P}^{-1}\mathbf{AP} = \boldsymbol{\Lambda}$. □

Outline

Art Transcends Physics?

Definition of Spectrum

Definition

The spectrum of a self-adjoint $n \times n$ matrix $\mathbf{A}$
is the set $\{\lambda_1, \lambda_2, \ldots, \lambda_m\}$ of its $m \leq n$ distinct eigenvalues.

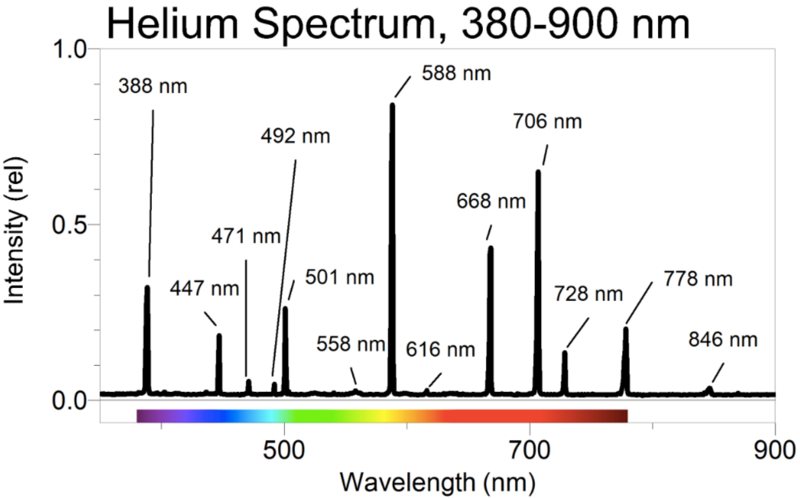
Isaac Newton introduced the word "spectrum" (ghosts?)
to describe the decomposition of sunlight into different colours
that we observe in rainbows,
and that he could produce using a prism.

These different colours were later identified with:

1. the wavelengths (ℓ) and corresponding frequencies (f)
   of different kinds of light, whose speed in a vacuum
   is $c = \ell f = 299,792,458$ metres per second
   (by the modern definition of a metre);
2. the eigenvalues of a self-adjoint matrix that appears
   in the Schrödinger wave equation of quantum mechanics.

Physicists used the spectrum illustrated on the next slide
to help discover the "new" element helium in the sun's atmosphere.

The Visible Part of the Helium Emission Spectrum



Note: nm is an abbreviation for nanometre $= 10^{-9}$ metre, which is one millionth of a millimetre.