Lecture Notes 1: Matrix Algebra
Part D: Similar Matrices and Diagonalization

Peter J. Hammond

revised 2018 September 21st

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

Definitions in the Real Case

Definition

Consider any $n \times n$ matrix $\mathbf{A}$.

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

In this case the solution $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$ 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.

The Eigenspace

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

Given any two eigenvectors $\mathbf{x}, \mathbf{y} \in E_\lambda$ and any two scalars $\alpha, \beta \in \mathbb{R}$, 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 $\mathbb{R}^n$ 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. □

Characteristic Equation

The equation $\mathbf{A}\mathbf{x} = \lambda\mathbf{x}$ holds for $\mathbf{x} \neq \mathbf{0}$
if and only if $\mathbf{x} \neq \mathbf{0}$ solves $(\mathbf{A} - \lambda\mathbf{I})\mathbf{x} = \mathbf{0}$.

This holds iff the matrix $\mathbf{A} - \lambda\mathbf{I}$ is singular,
which holds iff 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 $|\mathbf{A} - \lambda\mathbf{I}|$ of degree n.

Suppose $|\mathbf{A} - \lambda\mathbf{I}| = 0$ has k distinct real roots $\lambda_1, \lambda_2, \ldots, \lambda_k$
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 real 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 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 $\mathbf{D} = \mathbf{diag}(d_1, d_2, \ldots, d_n)$,
the characteristic equation $|\mathbf{D} - \lambda \mathbf{I}| = 0$
takes the degenerate form $\prod_{k=1}^{n}(d_k - \lambda) = 0$.

So the eigenvalues are the diagonal elements.

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

The kth vector $\mathbf{e}^k = (\delta_{jk})_{j=1}^{n}$
of the canonical orthonormal basis of $\mathbb{R}^n$
always solves the equation $\mathbf{Dx} = d_k \mathbf{x}$,
and so is an eigenvector associated with the eigenvalue d_k.

The Case of a Diagonal Matrix, II

Apart from non-zero multiples of $\mathbf{e}^k$,
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

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

Example with No Real Eigenvalues, I

Recall that a 2-dimensional *rotation matrix* takes the form

$$\mathbf{R}_\theta := \begin{pmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{pmatrix}$$

for $\theta \in \mathbb{R}$, which is the *angle of rotation* measured in radians.

The rotation $\mathbf{R}_\theta$ transforms any vector $\mathbf{x} = (x_1, x_2) \in \mathbb{R}^2$ to

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

Introduce polar coordinates (r, η),
where $\mathbf{x} = (x_1, x_2) = r(\cos\eta, \sin\eta)$. Then

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

This makes it easy to verify that $\mathbf{R}_{\theta+2k\pi} = \mathbf{R}_\theta$ for all $\theta \in \mathbb{R}$ and $k \in \mathbb{Z}$, and that $\mathbf{R}_\theta \mathbf{R}_\eta = \mathbf{R}_\eta \mathbf{R}_\theta = \mathbf{R}_{\theta+\eta}$ for all $\theta, \eta \in \mathbb{R}$.

Example with No Real Eigenvalues, II

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. There is a degenerate case when $\cos\theta = 1$ because $\theta = 2k\pi$ for some $k \in \mathbb{Z}$.

   Then $\mathbf{R}_\theta$ reduces to the identity matrix $\mathbf{I}_2$.
2. Otherwise, the real matrix $\mathbf{R}_\theta$ has no real eigenvalues.

   Indeed, if $\cos\theta < 1$, the characteristic equation has two roots $\lambda = \cos\theta \pm i \sin\theta = e^{\pm i\theta}$.

   Because $\sin\theta = \sqrt{1 - \cos^2\theta} \neq 0$, there are two distinct complex conjugate eigenvalues.

   The associated eigenspaces must both consist of complex eigenvectors.

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

Complex Eigenvalues

To consider complex eigenvalues properly,
we need to leave $\mathbb{R}^n$ and consider instead 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 the plane $\mathbb{C}$ of complex numbers,
rather than the line $\mathbb{R}$ of real numbers.

Suppose $\mathbf{A}$ is any $n \times n$ matrix
whose elements may be real or complex.

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,
in which case that solution $\mathbf{x} \in \mathbb{C}^n \setminus \{\mathbf{0}\}$ is an eigenvector.

Fundamental Theorem of Algebra

Theorem

Let $P(\lambda) = \lambda^n + \sum_{k=0}^{n-1} p_k \lambda^k$
be a polynomial function of λ *of degree* n *in the complex plane* $\mathbb{C}$*.*

Then there exists at least one root $\hat{\lambda} \in \mathbb{C}$ *such that* $P(\hat{\lambda}) = 0$*.*

Corollary

The polynomial $P(\lambda)$ *can be factorized*
as the product $P_n(\lambda) \equiv \prod_{r=1}^{n}(\lambda - \lambda_r)$ *of exactly* n *linear terms.*

Proof.

The proof will be by induction on n.

When $n = 1$ one has $P_1(\lambda) = \lambda + p_0$, whose only root is $\lambda = -p_0$.

Suppose the result is true when $n = m - 1$.

By the fundamental theorem of algebra,
there exists $\hat{\lambda} \in \mathbb{C}$ such that $P_m(\hat{\lambda}) = 0$.

Polynomial division gives $P_m(\lambda) \equiv P_{m-1}(\lambda)(\lambda - \hat{\lambda})$, etc. □

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 a polynomial of degree n.

By the fundamental theorem of algebra, together with its corollary, the polynomial $|\lambda \mathbf{I} - \mathbf{A}|$ equals the product $\prod_{r=1}^{n} (\lambda - \lambda_r)$ of n linear terms.

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

Eigenvalues and Eigenvectors

Real Case

The Complex Case

Eigenvectors are Linearly Independent

Diagonalizing a General Matrix

Similar Matrices

Diagonalizing a Symmetric Matrix

A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements

Rayleigh Quotient

The Spectral Theorem

Quadratic Forms and Their Definiteness

Quadratic Forms

The Eigenvalue Test of Definiteness

Eigenvectors Are 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 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

$$\alpha_k(\lambda_m - \lambda_k)\mathbf{x}_k = 0 \quad \text{for } k = 1, \dots, m-1$$

But we are assuming that $\lambda_m \not\in \{\lambda_k\}_{k=1}^{m-1}$,
so $\lambda_m - \lambda_k \neq 0$ for $k = 1, \dots, m-1$.

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

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 $\mathbf{x}_k \in E_{\lambda_k}$ $(k = 1, \dots, m-1)$. □

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

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

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

in which case we write $\mathbf{A} \sim \mathbf{B}$.

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

Then $\mathbf{x}$ solves $\mathbf{Ax} = \mathbf{SBS}^{-1}\mathbf{x} = \lambda\mathbf{x}$.

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

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 $\mathbf{x} = \mathbf{Sy} = \mathbf{0}$ 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 $\mathbf{B} = \mathbf{S}^{-1}\mathbf{SA}$,
then $(\lambda, \mathbf{Sy})$ is an eigenpair of $\mathbf{A}$. □

Diagonalization

Definition

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

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 Part 1

Suppose that $\mathbf{AS} = \mathbf{S\Lambda}$ 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 $i, k \in \{1, 2, \ldots, n\}$,
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 $\mathbf{S}$ must be invertible:

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

□

Proofs of 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: This follows from the general property that similar matrices have the same spectrum of eigenvalues.

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

Complex Conjugates and Adjoint Matrices

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

The complex conjugate of c is $\bar{c} = a - ib$.

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

Any $m \times n$ complex matrix $\mathbf{C} = (c_{ij})_{m \times n} \in \mathbb{C}^{m \times n}$ can be written as $\mathbf{A} + i\mathbf{B}$, where $\mathbf{A}$ and $\mathbf{B}$ are real $m \times n$ matrices.

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

This is the transpose of the matrix $\mathbf{A} - i\mathbf{B}$ whose elements are the complex conjugates $\bar{c}_{jk}$ of the corresponding elements of $\mathbf{C}$.

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

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 self-adjoint
just in case $\mathbf{C}^* = \mathbf{C}$, which 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.

Theorem

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

Proof that Eigenvalues are Real

Suppose that the scalar $\lambda \in \mathbb{C}$ and vector $\mathbf{x} \in \mathbb{C}^n$ together satisfy the eigenvalue equation $\mathbf{A}\mathbf{x} = \lambda \mathbf{x}$ for any $\mathbf{A} \in \mathbb{C}^{n \times n}$.

Taking complex conjugates throughout, one has $\bar{\lambda}\mathbf{x}^* = \mathbf{x}^*\mathbf{A}^*$.

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 $\mathbf{A}^* = \mathbf{A}$,
subtracting the second equation from the first gives

$$\mathbf{x}^*\mathbf{A}\mathbf{x} - \mathbf{x}^*\mathbf{A}^*\mathbf{x} = \mathbf{x}^*(\mathbf{A} - \mathbf{A}^*)\mathbf{x} = 0 = (\lambda - \bar{\lambda})(\mathbf{x}^*\mathbf{x})$$

But in case $\mathbf{x}$ is an eigenvector, one has $\mathbf{x} = (x_i)_{i=1}^n \in \mathbb{C}^n \setminus \{\mathbf{0}\}$
and so $\mathbf{x}^*\mathbf{x} = \sum_{i=1}^n |x_i|^2 > 0$.

Because $0 = (\lambda - \bar{\lambda})\mathbf{x}^*\mathbf{x}$, it follows that the eigenvalue λ
satisfies $\lambda - \bar{\lambda} = 0$, implying that λ is real. $\square$

Orthogonal Projections

Definition

An $n \times n$ matrix $\mathbf{P}$ is an orthogonal projection if $\mathbf{P}^2 = \mathbf{P}$
and $\mathbf{u}^\top \mathbf{v} = 0$ whenever $\mathbf{P}\mathbf{v} = \mathbf{0}$ and $\mathbf{u} = \mathbf{P}\mathbf{x}$ for some $\mathbf{x} \in \mathbb{R}^n$.

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 a symmetric orthogonal projection onto L.*
2. *The matrix $\mathbf{I} - \mathbf{P}$ is a symmetric 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

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

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

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

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

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

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

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

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

so $\mathbf{y} \in L^\perp$. □

Proof of Part 3

For any vector $\mathbf{v} = \mathbf{X}\mathbf{b} \in L$ and $\mathbf{y} \in \mathbb{R}^n$,
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 $\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{e}$).

Also, define $\hat{\mathbf{v}} := \mathbf{X}\hat{\mathbf{b}} = \mathbf{P}\mathbf{y}$. Because $\mathbf{P}^\top \mathbf{P} = \mathbf{P}^\top = \mathbf{P} = \mathbf{P}^2$,

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

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 $\mathbf{v} = \hat{\mathbf{v}}$. $\square$

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

A Trick Function for Generating Eigenvalues

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 $\mathbf{x} = \mathbf{0}$.

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

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

At any stationary point $\hat{\mathbf{x}} \neq \mathbf{0}$ where $\partial f / \partial x_h = 0$ for all h,
one therefore has $(\hat{\mathbf{x}}^\top \hat{\mathbf{x}})\mathbf{A}\hat{\mathbf{x}} = (\hat{\mathbf{x}}^\top \mathbf{A}\hat{\mathbf{x}})\hat{\mathbf{x}}$
and so $\mathbf{A}\hat{\mathbf{x}} = \lambda \hat{\mathbf{x}}$ where $\lambda = f(\hat{\mathbf{x}})$.

That is, a stationary point $\hat{\mathbf{x}} \neq \mathbf{0}$ must be an eigenvector,
with the corresponding function value $f(\hat{\mathbf{x}})$
as the associated eigenvalue.

More Properties of the Rayleigh Quotient

Using the Rayleigh quotient

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

one can state and prove the following lemma.

Lemma

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

1. *has a maximum eigenvalue λ^* at an eigenvector $\mathbf{x}^*$ where f attains its maximum;*
2. *has a minimum eigenvalue λ_* at an eigenvector $\mathbf{x}_*$ where f attains its minimum;*
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 compact subset of $\mathbb{R}^n$,
and the Rayleigh quotient function f restricted to S^{n-1} is continuous.

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 stationary at any maximum point $\mathbf{x}^*$,
as well as at any minimum point $\mathbf{x}_*$.

But stationary points must be eigenvectors.

This proves parts 1 and 2 of the lemma. □

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

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 vectors,
as well as the columns of an $n \times m$ *matrix* $\mathbf{U}$.

Then there is at least one more eigenvector $\mathbf{x}$
that satisfies $\mathbf{U}^\top \mathbf{x} = 0$
— 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 $\mathbf{AU} = \mathbf{U\Lambda}$ where $\mathbf{\Lambda} := \mathbf{diag}(\lambda_k)_{k=1}^n$.

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 $\hat{\mathbf{A}} := (\mathbf{I} - \mathbf{U}\mathbf{U}^\top)\mathbf{A}(\mathbf{I} - \mathbf{U}\mathbf{U}^\top)$, which is symmetric because both $\mathbf{A}$ and $\mathbf{U}\mathbf{U}^\top$ are symmetric.

Note that

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

This matrix $\hat{\mathbf{A}}$ has at least one eigenvalue λ, which must be real, and an associated eigenvector $\mathbf{x} \neq \mathbf{0}$, which together satisfy

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

Pre-multiplying each side of the last equation by $\mathbf{U}^\top$ shows that

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

This leaves two cases, to be considered on the next two slides.

Constructive Proof, Part 3

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 satisfies $\mathbf{U}^\top \mathbf{x} = \mathbf{0}_m^\top$.

But then the earlier equation

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

implies that

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

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 $\lambda = 0$,
implying that $\hat{\mathbf{A}} = \mathbf{0}$ and so $\mathbf{A} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top$.

Then any vector $\mathbf{x} \neq \mathbf{0}$ satisfying $\mathbf{U}^\top \mathbf{x} = \mathbf{0}$
must satisfy $\mathbf{A}\mathbf{x} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^\top \mathbf{x} = \mathbf{0}$,
implying that $\mathbf{x}$ is an eigenvector of $\mathbf{A}$.

In both cases there is an eigenvector $\mathbf{x}$ of $\mathbf{A}$ satisfying $\mathbf{U}^\top \mathbf{x} = \mathbf{0}_m^\top$.

Spectral Theorem

Theorem

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

1. *its eigenvectors span 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 $\mathbf{\Lambda}$;*

Proof of Spectral Theorem, Part 1

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

The associated eigenvector $\mathbf{x}$, normalized to satisfy $\mathbf{x}^\top \mathbf{x} = 1$, forms an orthonormal set $\{\mathbf{u}_1\}$.

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

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

The "useful lemma" shows that, if the hypothesis holds
for any $m = 1, 2, \ldots, n - 1$, 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, taking $\mathbf{P}$ as an orthogonal matrix whose columns are an orthonormal set of n eigenvectors, we obtain $\mathbf{P}^\top \mathbf{A}\mathbf{P} = \mathbf{P}^{-1}\mathbf{A}\mathbf{P} = \mathbf{\Lambda}$.

Outline

Eigenvalues and Eigenvectors

Real Case

The Complex Case

Eigenvectors are Linearly Independent

Diagonalizing a General Matrix

Similar Matrices

Diagonalizing a Symmetric Matrix

A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements

Rayleigh Quotient

The Spectral Theorem

Quadratic Forms and Their Definiteness

Quadratic Forms

The Eigenvalue Test of Definiteness

Definition of Quadratic Form

Definition

A quadratic form on the n-dimensional Euclidean space R^n is a mapping

$$R^n \ni \mathbf{x} \mapsto \mathbf{x}^\top \mathbf{Q} \mathbf{x} = \sum_{i=1}^{n} \sum_{j=1}^{n} x_i q_{ij} x_j \in \mathbb{R}$$

where $\mathbf{Q}$ is a symmetric $n \times n$ matrix.

The quadratic form $\mathbf{x}^\top \mathbf{Q} \mathbf{x}$ is diagonal just in case the matrix $\mathbf{Q}$ is diagonal, with $\mathbf{Q} = \mathbf{\Lambda} = \mathbf{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$.

In this case $\mathbf{x}^\top \mathbf{Q} \mathbf{x}$ reduces to $\mathbf{x}^\top \mathbf{\Lambda} \mathbf{x} = \sum_{i=1}^{n} \lambda_i \, (x_i)^2$.

Symmetry Loses No Generality

Requiring $\mathbf{Q}$ to be symmetric loses no generality.

This is because, given a general non-symmetric $n \times n$ matrix $\mathbf{A}$, repeated transposition implies that

$$\mathbf{x}^\top \mathbf{A} \mathbf{x} = (\mathbf{x}^\top \mathbf{A} \mathbf{x})^\top = \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} = \tfrac{1}{2} \mathbf{x}^\top (\mathbf{A} + \mathbf{A}^\top) \mathbf{x}$$

Hence $\mathbf{x}^\top \mathbf{A} \mathbf{x} = \mathbf{x}^\top \mathbf{A}^\top \mathbf{x} = \mathbf{x}^\top \mathbf{Q} \mathbf{x}$
where $\mathbf{Q}$ is the symmetrized matrix $\frac{1}{2}(\mathbf{A} + \mathbf{A}^\top)$.

Note that $\frac{1}{2}(\mathbf{A} + \mathbf{A}^\top)$ is indeed symmetric;
indeed, the definition of transpose implies that

$$(\mathbf{Q}^\top)_{ij} = (\mathbf{Q})_{ji} = \tfrac{1}{2}[(\mathbf{A})_{ji} + (\mathbf{A}^\top)_{ji}] = \tfrac{1}{2}[(\mathbf{A}^\top)_{ij} + (\mathbf{A})_{ij}] = (\mathbf{Q})_{ij}$$

for all (i, j).

Definiteness of a Quadratic Form

When $\mathbf{x} = \mathbf{0}$, then $\mathbf{x}^\top \mathbf{Q} \mathbf{x} = 0$. Otherwise:

Definition

The quadratic form $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{x}^\top \mathbf{Q} \mathbf{x} \in \mathbb{R}$ is:

positive definite just in case $\mathbf{x}^\top \mathbf{Q} \mathbf{x} > 0$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;

negative definite just in case $\mathbf{x}^\top \mathbf{Q} \mathbf{x} < 0$ for all $\mathbf{x} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$;

positive semi-definite just in case $\mathbf{x}^\top \mathbf{Q} \mathbf{x} \geq 0$ for all $\mathbf{x} \in \mathbb{R}^n$;

negative semi-definite just in case $\mathbf{x}^\top \mathbf{Q} \mathbf{x} \leq 0$ for all $\mathbf{x} \in \mathbb{R}^n$;

indefinite just in case there exist $\mathbf{x}^+$ and $\mathbf{x}^-$ in $\mathbb{R}^n$ such that $(\mathbf{x}^+)^\top \mathbf{Q} \mathbf{x} > 0$ and $(\mathbf{x}^-)^\top \mathbf{Q} \mathbf{x} < 0$.

Definiteness of a Diagonal Quadratic Form

Theorem

The diagonal quadratic form $\sum_{i=1}^{n} \lambda_i (x_i)^2 \in \mathbb{R}$ *is:*

positive definite *if and only if* $\lambda_i > 0$ *for* $i = 1, 2, \ldots, n$;

negative definite *if and only if* $\lambda_i < 0$ *for* $i = 1, 2, \ldots, n$;

positive semi-definite *if and only if* $\lambda_i \geq 0$ *for* $i = 1, 2, \ldots, n$;

negative semi-definite *if and only if* $\lambda_i \leq 0$ *for* $i = 1, 2, \ldots, n$;

indefinite *if and only if there exist* $i, j \in \{1, 2, \ldots, n\}$ *such that* $\lambda_i > 0$ *and* $\lambda_j < 0$.

Proof.

The proof is left as an exercise.

The result is obvious if $n = 1$, and straightforward if $n = 2$.

Working out these two cases first suggests the proof for $n > 2$. □

Outline

Eigenvalues and Eigenvectors
Real Case
The Complex Case
Eigenvectors are Linearly Independent

Diagonalizing a General Matrix
Similar Matrices

Diagonalizing a Symmetric Matrix
A Symmetric Matrix has only Real Eigenvalues

Orthogonal Projections and Complements
Rayleigh Quotient
The Spectral Theorem

Quadratic Forms and Their Definiteness
Quadratic Forms
The Eigenvalue Test of Definiteness

Diagonalizing Quadratic Forms

Consider a quadratic form $\mathbb{R}^n \ni \mathbf{x} \mapsto \mathbf{x}^\top \mathbf{Q} \mathbf{x} \in \mathbb{R}$ where,
without losing generality,
we assume that the $n \times n$ matrix $\mathbf{Q}$ is symmetric.

By the spectral theorem for symmetric matrices,
there exists a matrix $\mathbf{P}$ that diagonalizes $\mathbf{Q}$,
meaning that $\mathbf{P}^{-1}\mathbf{Q}\mathbf{P}$ is a diagonal matrix that we denote by $\mathbf{\Lambda}$.

Moreover $\mathbf{P}$ can be made orthogonal, meaning that $\mathbf{P}^{-1} = \mathbf{P}^\top$.

Given any $\mathbf{x} \neq \mathbf{0}$, because $\mathbf{P}^{-1}$ exists, we can define $\mathbf{y} = \mathbf{P}^{-\mathbf{1}}\mathbf{x}$.

This implies that $\mathbf{x} = \mathbf{P}\mathbf{y}$, where $\mathbf{y} \neq \mathbf{0}$ because $(\mathbf{P}^{-\mathbf{1}})^{-\mathbf{1}} = \mathbf{P}$.

Then $\mathbf{x}^\top \mathbf{Q} \mathbf{x} = \mathbf{y}^\top \mathbf{P}^\top \mathbf{Q} \mathbf{P} \mathbf{y} = \mathbf{y}^\top \mathbf{\Lambda} \mathbf{y}$,
so the diagonalization leads to a diagonal quadratic form.

The Eigenvalue Test

A standard result says that $\mathbf{Q}$
and its diagonalization $\mathbf{P}^{-1}\mathbf{Q}\mathbf{P}$ have the same set of eigenvalues.

From the theorem on the definiteness of a diagonal quadratic form,
it follows that:

Theorem

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

positive definite *if and only if all its eigenvalues are positive;*

negative definite *if and only if all its eigenvalues are negative;*

positive semi-definite *if and only if all its eigenvalues are non-negative;*

negative semi-definite *if and only if all its eigenvalues are non-positive;*

indefinite *if and only if it has both positive and negative eigenvalues.*